Baysian HW 3

Brandon Lucio

10/16/2021

Problem 1

Glass fragments found on a suspect's shoes or clothes are often used to connect the suspect to a crime scene. The index of refraction of the fragments are compared to the refractive index of the glass from the crime scene. To make this comparison rigorous, we need to know the variability the index of refraction is over a pane of glass. Bennett et al. (2003) analyzed the refractive index in a pane of float glass, searching for any spatial pattern. Here are samples of the refractive index from the edge and from the middle of the pane.

```r
# Placing sample data into different vectors
Edge_Pane <- c(1.51996,1.51997,1.51998,1.52000,1.51998,1.52004,1.52,1.52001,1.52,1.51997)

Middle_Pane <- c(1.52001,1.51999,1.52004,1.51997,1.52005,1.52,1.52004,1.52002,1.52004,1.51996)
```

Part A

Suppose glass at the edge of the pane is $normal(\mu_1, \sigma_1^2)$, where $\sigma_1 = .00003$. Calculate the posterior distribution of μ_1 when you use a $normal(1.52, .0001^2)$ prior for μ_1. Show all work.

1

a) $A|\mu, \sim N(\mu, \sigma^2)$ where $\sigma = 0.00003$

Let $p(\mu) \sim N(1.52, 0.0001^2)$

Now we try to generalize this normal sample with normal prior

using the given info we get

plug distributions $p(\mu|A) \propto p(\mu)p(A|\mu)$

remove constants

$$\propto \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right] \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x_i-\mu)^2}{2\sigma^2}\right]$$

$$\propto \frac{1}{\sqrt{\sigma_0^2\sigma^{2n}}} \exp\left[\frac{-\mu^2+2\mu\mu_0-\mu_0^2}{2\sigma_0^2} - \sum_{i=1}^{n} \frac{x_i^2 - 2\mu x_i + \mu^2}{2\sigma^2}\right]$$

get common denominator

$$\propto \exp\left[\frac{-\mu^2+2\mu\mu_0-\mu_0^2}{2\sigma_0^2} - \frac{\Sigma x_i^2 - 2n\bar{x}\mu + \mu}{2\sigma^2}\right]$$

devide by $\{\sigma^2 + n\sigma_0^2\}$ and remove all constant terms (not μ)

$$\propto \exp\left[\frac{-\mu^2 + 2\mu\frac{\mu_0\sigma^2+n\bar{x}\sigma_0^2}{\sigma^2+n\sigma_0^2} - \left(\frac{\mu_0\sigma^2+n\bar{x}\sigma_0^2}{\sigma^2+n\sigma_0^2}\right)^2}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}\right]$$

$$\propto \exp\left[-\frac{\left(\mu - \frac{\mu_0\sigma^2+n\bar{x}\sigma_0^2}{\sigma^2+n\sigma_0^2}\right)^2}{2\left(\sigma_0^2\sigma^2/\sigma^2+n\sigma_0^2\right)}\right]$$ this is normal kernel

thus

$$\mu|A \sim N(\mu^*, \sigma^{2*})$$

where

$$\sigma^{2*} = \frac{\sigma^2\sigma_0^2}{\sigma^2+n\sigma_0^2}$$ and

$$\mu^* = \frac{\mu_0\sigma^2 + n\bar{x}\sigma_0^2}{\sigma^2+n\sigma_0^2}$$

```r
# Using the derivation above, we create a function to make calculation easier.
#
# Functions to calculate posterior distribution mean and standard deviation
# with a Normal Sample with known sigma and a normal prior
#

post_mu <- function(sample, prior_mu, prior_sig, sample_sig){
  n <- length(sample)
  xbar <- mean(sample)

  mu <- (prior_mu*sample_sig^2 + n*xbar*prior_sig^2)/(sample_sig^2 + n*prior_sig^2)

  return(mu)
}

post_sigma <- function(sample, prior_mu, prior_sig, sample_sig){
  n <- length(sample)
```

```
  var <- (sample_sig^2*prior_sig^2)/(sample_sig^2 + n*prior_sig^2)

  return(sqrt(var))
}
```

```
# Given values
sigma <- 0.00003
sigma_0 <- 0.0001
mu_0 <- 1.52000

mu1_star <- post_mu(Edge_Pane,mu_0,sigma_0, sigma)
sigma1_star <-post_sigma(Edge_Pane,mu_0,sigma_0, sigma)
```

- Now we have $\mu_1|X \sim N(\mu^*, (\sigma^*)^2)$ where $\mu^* = 1.5199911$ and $\sigma^* = 9.4444283 \times 10^{-6}$

Part B

Suppose glass at the middle of the pane is $normal(\mu_2, \sigma_2^2)$, where $\sigma_2 = .00003$. Calculate the posterior distribution of μ_2 when you use a $normal(1.52, .0001^2)$ prior for μ_2. Show all work.

```
sigma <- 0.00003
sigma_0 <- 0.0001
mu_0 <- 1.52000

mu2_star <- post_mu(Middle_Pane,mu_0,sigma_0, sigma)
sigma2_star <-post_sigma(Middle_Pane,mu_0,sigma_0, sigma)
```

Part C

Find the posterior distribution of $\mu_d = \mu_1 - \mu_2$.

Part D

Find a 95% credible interval for μ_d above.

Part E

Based on part d), Perform a Bayesian test of the hypothesis: $H_0 : \mu_d = 0$ vs $H_1 : \mu_d \neq 0$

Problem 2

Suppose you wish to compare a new method of teaching to slow learners with the current standard method. You decide to base your comparison on the results of a reading test given at the end of a learning period of six months. Of a random sample of 22 slow learners, 10 are taught by the new method and 12 by the standard method. All 22 children are taught by qualified instructors under similar conditions for the designated six-month period. The results of the reading test at the end of this period are given below (assume that the assumptions stated above are satisfied):

Test Scores

new_method	standard_method
80	79
76	73
70	72
80	62
66	76
85	68
79	70
71	86
81	75
76	68
	73
	66

Part A

Use the t-test that was discuss in the class to test whether there exists the true mean difference between the test scores using an $\alpha = 0.05$ significance level.

```
t.test(new_method, standard_method)
```

```
##
##  Welch Two Sample t-test
##
## data:  new_method and standard_method
## t = 1.5643, df = 19.769, p-value = 0.1336
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.360091  9.493424
## sample estimates:
## mean of x mean of y
##  76.40000  72.33333
```

Part B

Use the Bayesian procedures under the non-informative priors to answer the following problems

i.

Are the test scores for the new method larger than the test scores for the standard method?

ii.

Is the variance of the test scores for the new method smaller than that for the standard method?

iii.

What are the posterior distributions of the coefficient of variation of each method?

iV.

What is the probability that a randomly selected learner taught by the new method will have better test scores than a randomly selected learner taught by the standard method?

Problem 3

Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ such that X_1 and X_2 are independent. Define $Y = X_1 + X_2$. The goal of this problem is to determine the distribution of Y; i.e., the sum of two independent normal random variables.

Part A

Obviously, this is trivial, so simply state the distribution of Y.

- Here $Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Part B

Approach 1: Consider using Monte Carlo sampling to obtain a histogram and kernel density estimate (see the code that I have provided) of the pdf of Y by directly sampling both X_1 and X_2. Over plot the true density of Y and comment. Note, you should make use of large enough Monte Carlo sample that your results are reasonable.

Part C

Approach 2: Note, the distribution of Y can also be obtained through the convolution of the probability distributions of X_1 and X_2. Sketch out theoretically how this would be done. Based on this idea, create a Monte Carlo sampling technique which can be used to approximate the pdf of Y evaluated at any point in the support. Use this function and add the approximation based on this technique to the Figure described in part (b) above.