# HW3 R Markdown

Santanu Mukherjee

10/24/2021

# R Markdown

## Problem 1

**Answer 1a**

Name → Santanu Mukherjee Zes254

Problem 1

(a) $X \sim N(\mu_1, \sigma_1^2)$, $\sigma_1 = 0.00003$ $\sum x_i = 15.19991$

Prior of $\mu_1$ follows $N(\underbrace{1.52000}_{\mu_k}, \underbrace{0.0001^2}_{\sigma_k^2})$

Posterior distribution.

$$P(\mu_1 | x) = P(x|\mu_1) \cdot P(\mu_1)$$

$$P(\mu_1|x) \propto \exp\left\{-\frac{1}{2\sigma_1^2}\sum(x_i-\mu_1)^2\right\} \exp\left\{-\frac{1}{2\sigma_k^2}(\mu_1-\mu_k)^2\right\}$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_1^2} n\mu_1^2 - \frac{2}{\sigma_1^2}\sum x_i \mu_1 + \frac{1}{\sigma_k^2}\mu_1^2 - \frac{2}{\sigma_k^2}\mu_k \mu_1\right)\right]$$

Putting values for $\mu_k = 1.52000$, $\sigma_k^2 = 0.0001^2$, $\sigma_1 = 0.00003$, $n = 10$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{10}{0.00003^2} + \frac{1}{0.0001^2}\right)\mu_1^2 - \left(\frac{2}{0.00003^2} \times 15.19991 + \frac{2}{0.0001^2} \times 1.52000\right)\mu_1\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\frac{1}{8.92e\text{-}11}\left[\mu_1 - \frac{17040788889}{1121111111}\right]^2\right\}$$

$$\downarrow \quad 8.92e\text{-}11 \qquad \downarrow \quad 1.51999108$$

$\therefore \mu_1|x \sim N(1.51999108, 8.92e\text{-}11)$

means $\mu_1|x \sim N(\underbrace{1.51999108}_{\mu}, \underbrace{0.0000000000891972}_{\sigma^2})$

Problem-1-a

**Answer 1 b and 1 c**

Problem 1

(b) $X \sim N(\mu_2, \sigma_2^2)$, $\sigma_2 = 0.00003$, $n = 10$, $\sum x_i = 15.20012$

So, based on similar derivation. $P(\mu_2) \sim N(\underbrace{1.52000}_{\mu_k}, \underbrace{0.0001^2}_{\sigma_k^2})$

$$P(\mu_2|x) = P(x|\mu_2)\, P(\mu_2)$$

$$P(\mu_2|x) \propto \exp\left\{-\frac{1}{2\sigma_2^2}\sum(x_i-\mu_2)^2\right\}\exp\left\{-\frac{1}{2\sigma_k^2}(\mu_2-\mu_k)^2\right\}$$

Putting the values:

$$P(\mu_2|x) \propto \exp\left\{-\frac{1}{2}\underbrace{\frac{1}{8.92e\text{-}11}}_{8.92e\text{-}11}\left[\mu_1 - \underbrace{\frac{17041022222}{11211111111}}_{1.520011893}\right]^2\right\}$$

So: $\mu_2|x \sim N(\underbrace{1.520011893}_{\mu}, \underbrace{8.92e\text{-}11}_{\sigma^2})$

Problem 1

(c) Assuming edge of panel and middle of pane are independently distributed,

$\mu_d = \mu_1 - \mu_2$ is normally distributed

with mean $= \mu_1 - \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$

mean of $\mu_d = -0.0000208127$, $\sigma^2 = 1.78394e\text{-}10$

So, $\mu_d \sim N(-0.0000208127, 1.78394e\text{-}10)$

Problem-1-b-c

### Answer 1d

The 95% credible interval for $\mu_d$ is

```
qnorm(c(0.025, 0.975), -0.0000208127, 0.0000133564)
```

```
## [1] -4.699076e-05  5.365363e-06
```

### Answer 1e

Here the null hypothesis is $H_0$ : $\mu_d$ = 0. Now looking at the 95% credible interval of $\mu_d$, as we can see that the value of $\mu_d$ = 0 falls in between the 95% credible interval, and so we can say that we FAIL to REJECT $H_0$.

### Answer 2a

Zes 254

Problem 2
(a) Let $X_1$ = new method, $X_2$ = Standard method

Now, the t-test has 2 different Scenarios

Scenario ① Equal variance $\sigma_1^2 = \sigma_2^2$

In this Case $H_0: \mu_1 = \mu_2$ vs Alternative $H_0: \mu_1 \neq \mu_2$

Test Statistic $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ ; $S_c = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

where $s_1$ and $s_2$ are sample standard deviations.

Scenario ② Variance NOT EQUAL, $\sigma_1^2 \neq \sigma_2^2$

Test Statistic $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{\min(n_1-1,\, n_2-1)}$

So, in this problem we do not know whether the variances are equal, so using Scenario 2.

from the sample, $n_1 = 10$, $n_2 = 10$, $\bar{X}_1 = 76.4$, $S_1 = 5.83476$

$\bar{X}_2 = 72.33$, $S_2 = 6.34369$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = 1.627776 \sim t_9$$

Now for $\alpha = 0.05$ the critical value is 1.833

So, we can say that if we compare the 2 test scores, there exists true mean difference at $\alpha = 0.05$ Significance level.

Problem-2-a

### Answer 2b

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

### Answer 2b i

```
## [1] "P(Test Scores of new method > Test Scores for Standard Method) : 0.93333"
```

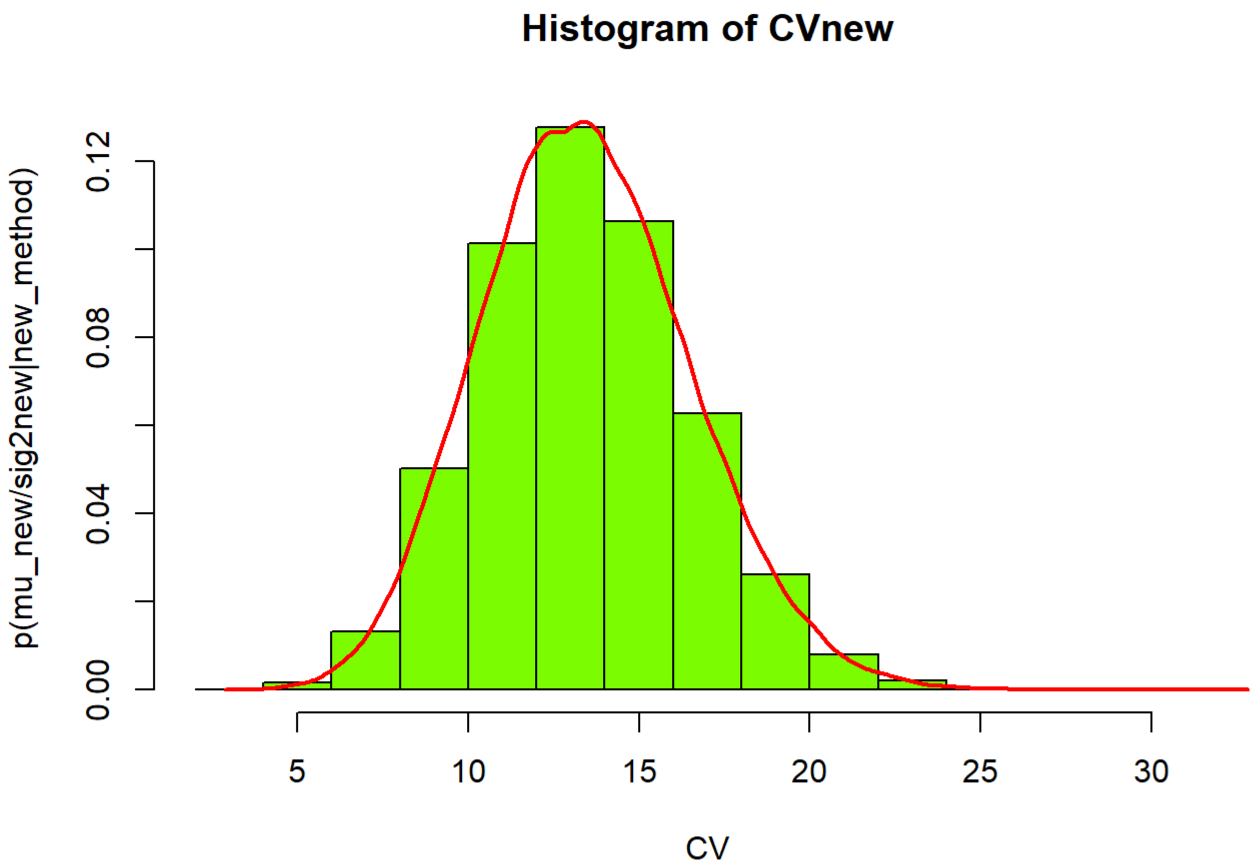
### Answer 2b ii

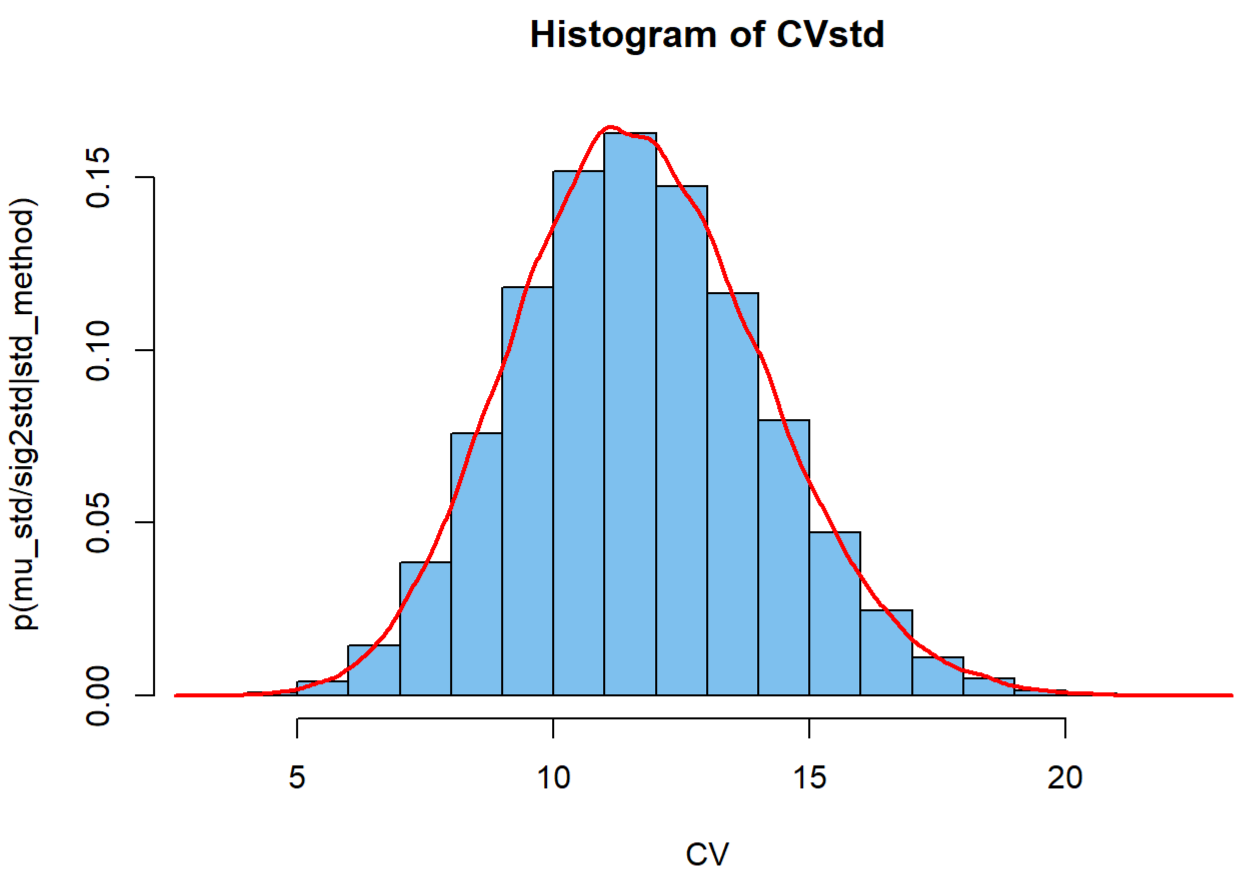
Is the variance of the test scores for the new method smaller than that for the standard method ?

```
## [1] "P(variance of test scores for new method < Variance of test scores for Standard Method)
: 0.60688"
```

So, the answer is YES

### Answer 2b iii


Histogram of CVnew
p(mu_new/sig2new|new_method)
0.00
0.04
0.08
0.12
5
10
15
20
25
30
CV


Histogram of CVstd
p(mu_std/sig2std|std_method)
0.00
0.05
0.10
0.15
5
10
15
20
CV

### Answer 2b iv

```
## [1] "Probability that a randomly selected learner taught by the new method will have better t
est scores than a randomly selected learner taught by the standard model is : 0.67266"
```

### Answer 3a

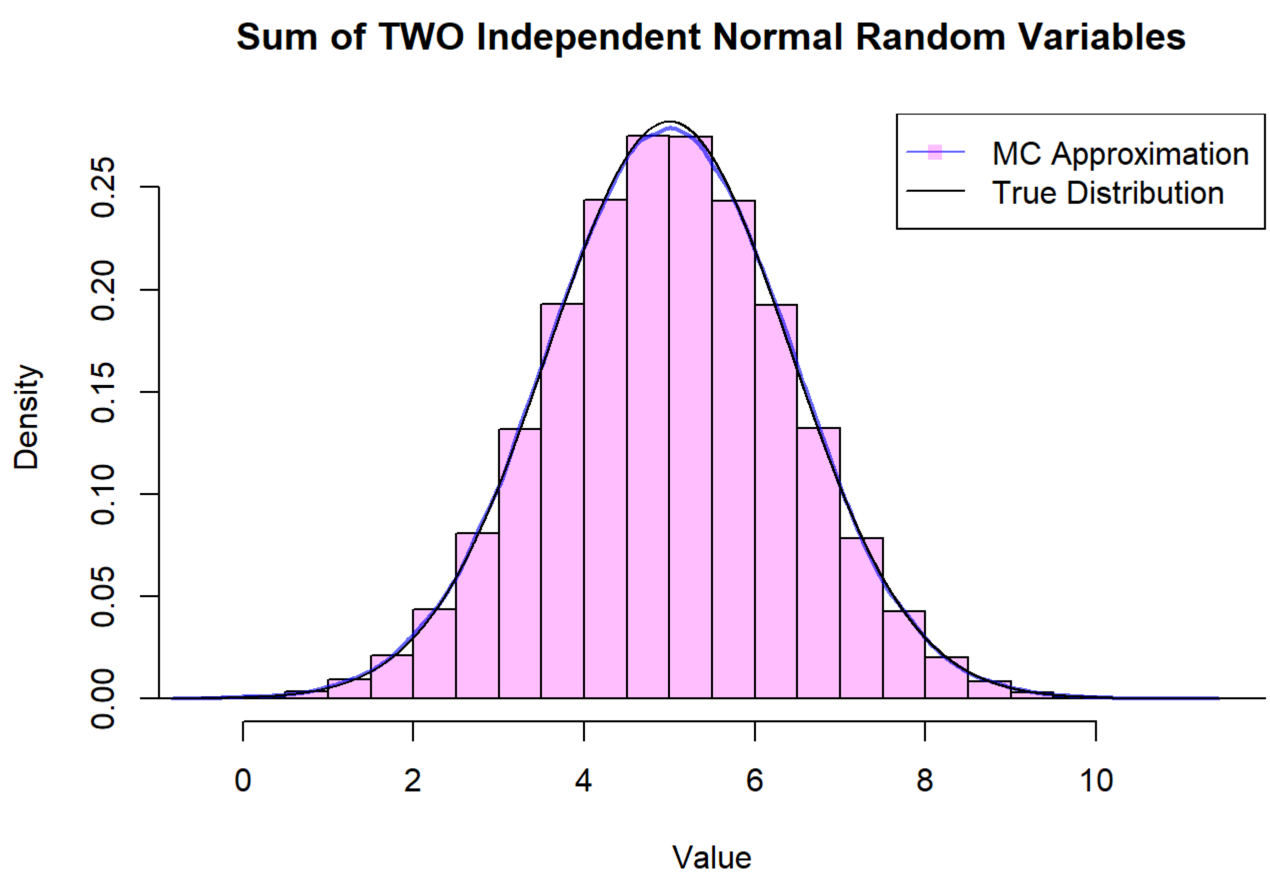
If $X_1$ ~ N($\mu_1$,$\sigma_1^2$) and $X_2$ ~ N($\mu_2$,$\sigma_2^2$) such that $X_1$ and $X_2$ are independent, then the distribution of $Y$ = $X_1$ + $X_2$ is also normally distributed.

So, $Y$ ~ N($\mu_1$ + $\mu_2$, $\sigma_1^2$ + $\sigma_2^2$).

### Answer 3b

Approach 1 - Using Monte Carlo sampling. Direct Sampling of $X_1$ and $X_2$.

I have taken $X_1$ ~ $N(2, 1)$ and $X_1$ ~ $N(3, 1)$. So ,this means $Y$ should be $N(5, 2)$, the above histogram validates this as it is centered around $5$ (which is mean of $Y$) and looking at the Empirical Rule, it can be said that the majority of the data lies within $3\sqrt{2}$, which means $3$ times standard deviation.
According to what I see, the generated distribution (by Monte Carlo method) is very close to the TRUE distribution.




### Answer 3c

Approach 2

For any specified $y$, we can approximate $P(y)$ by sampling $g$ from N($\mu_1$,$\sigma_1^2$) and then calculating the average density of $y - g$ when sampled from N($\mu_2$,$\sigma_2^2$). The GREEN line in the plot below has been generated in this manner.

Now, for each value of $y$, $100,000$ samples ($x_{1,i}$) were taken from N($\mu_1$,$\sigma_1^2$). Then the average density of $y - x_{1,i}$ in N($\mu_2$,$\sigma_2^2$) was used to estimate $P(y)$. The result of this (in GREEN) is plotted on top of the previous results tha you see above in part (b).