BIOS668 HW9
Sara O'Brien
4/18/23

<u>Honor Code:</u> On my honor, I have neither given nor received unauthorized aid on this assignment." *Sara O'Brien*

**Q1.**

| **Model** | **Prior** $P(\theta)$ | **Likelihood** | **Prior x Likelihood** | **Posterior** | **Model x Posterior** |
|---|---|---|---|---|---|
| 0 | 0.001 | 0 | 0 | 0 | 0 |
| 0.1 | 0.001 | 7.29e-19 | 7.29e-22 | 8.93e-16 | 8.93e-17 |
| 0.2 | 0.001 | 1.34e-13 | 1.34e-16 | 1.64e-10 | 3.28e-11 |
| 0.3 | 0.001 | 1.33e-10 | 1.33e-13 | 1.63e-7 | 4.89e-8 |
| 0.4 | 0.001 | 1.48e-8 | 1.48e-11 | 1.81e-5 | 7.24e-6 |
| 0.5 | 0.99 | 4.77e-7 | 4.72e-7 | 0.578 | 0.289 |
| 0.6 | 0.001 | 6.50e-6 | 6.50e-9 | 0.008 | 0.005 |
| 0.7 | 0.001 | 4.40e-5 | 4.40e-8 | 0.054 | 0.038 |
| 0.8 | 0.001 | 1.44e-4 | 1.44e-7 | 0.176 | 0.141 |
| 0.9 | 0.001 | 1.50e-4 | 1.50e-7 | 0.184 | 0.166 |
| 1.0 | 0.001 | 0 | 0 | 0 | 0 |
| Sum | 1.0 | | 8.17e-7 | 1 | 0.639 |

Proportion of B is 18/21 = .857, proportion of A is 3/21 = 0.143
$P(Y|\theta)$ = Likelihood: $\theta^{18} * (1-\theta)^3$
$P(\theta|Y)$ = Posterior: $P(\theta = \theta_i|Data) = (Prior * Likelihood)/Sum$ or P(Y)
P(Y) = Sum(P(Y, $\theta$) = Sum(P(Y| $\theta$)$P(\theta)$)

**Q2.**

Suppose that $x_1, \ldots, x_n$ are i.i.d $N(\theta, 1)$ and the prior distribution for $\theta$ is $\pi(\theta) \propto 1$.
The Bayesian predictive distribution of a future observation $z$ is defined as $p(z|x)$.

$$p(z|x) = \frac{p(z,x)}{p(x)}$$

$$p(z|x) = \int_\theta P(z|\theta) \cdot P(\theta|x)\, d\theta$$

$$P(z|\theta) \propto \int \exp\{-\tfrac{1}{2}(z-\theta)^2\}$$

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

$$\propto P(x|\theta)P(\theta)$$

$$\propto \cancel{z}\left[\frac{-1}{2}(x_i-\theta)^2\right] \cdot 1$$

$$\propto \exp\left\{\frac{-n}{2}(\theta-\bar{x})^2\right\} \sim N(\bar{x}, \tfrac{1}{2})$$

So, combining these we get

$$\int_\theta P(z|\theta) \cdot P(\theta|x)\, d\theta$$

$$= \int \exp\{-\tfrac{1}{2}(z-\theta)^2\} \cdot \exp\{-\tfrac{n}{2}(\theta-\bar{x})^2\}\, d\theta$$

$$= z|x \sim N(\bar{x}, 1+\tfrac{1}{n})$$

# Q3.

# R Code:

````
```{r, fig.width=5,fig.height=10}
# Set seed as PID
set.seed(730317945)

# Initialize matrix to store coefficients
coefficients_mat <- matrix(0, nrow = 300, ncol = 12)

# Replicate simulation 300 times
for (i in 1:300) {

  # Simulate data with covariates x1, x2, and z
  .n <- 1000
  .d <- data.frame(x1 = rnorm(.n))
  .d$x2 <- sqrt(0.5)*.d$x1 + rnorm(.n, sd=sqrt(0.5))
  .d$z <- as.numeric(.d$x1 + .d$x2 > 0)

  # Generate outcome
  .d$y <- 2.0 + 1.0*.d$x1 + 1.0*.d$x2 - 1.0*.d$z + rnorm(.n)

  # Generate error-prone covariates w1 and w2
  Sigma_error <- diag(c(0.20, 0.30))
  dimnames(Sigma_error) <- list(c("w1","w2"), c("w1","w2"))
  .d$w1 <- .d$x1 + rnorm(.n, sd = sqrt(Sigma_error["w1","w1"]))
  .d$w2 <- .d$x2 + rnorm(.n, sd = sqrt(Sigma_error["w2","w2"]))

  # Fit models without measurement error in covariates
  .mod0 <- lm(y ~ w1 + w2 + z, data = .d)

  # Fit model with measurement error in covariates
  .mod1 <- lm(y ~ x1 + x2 + z, data = .d)

  # Store coefficients
  coefficients_mat[i, ] <- cbind(c(2.0,1.0,1.0,-1.0), .mod1$coefficients, .mod0$coefficients)
}


colnames(coefficients_mat) <- c("true_intercept", "true_x1", "true_x2", "true_z", "intercept_ME", "x1_ME", "x2_ME", "z_ME", "intercept_noME", "x1_noME", "x2_noME", "z_noME")
analysis <- data.frame(coefficients_mat)

# Compute mean and standard deviation of coefficients
mean_intercept_ME <- mean(analysis$intercept_ME)
sd_intercept_ME <- sd(analysis$intercept_ME)
mean_intercept_noME <- mean(analysis$intercept_noME)
sd_intercept_noME <- sd(analysis$intercept_noME)

mean_x1_ME <- mean(analysis$x1_ME)
sd_x1_ME <- sd(analysis$x1_ME)
mean_x1_noME <- mean(analysis$x1_noME)
sd_x1_noME <- sd(analysis$x1_noME)

mean_x2_ME <- mean(analysis$x2_ME)
sd_x2_ME <- sd(analysis$x2_ME)
mean_x2_noME <- mean(analysis$x2_noME)
sd_x2_noME <- sd(analysis$x2_noME)

mean_z_ME <- mean(analysis$z_ME)
sd_z_ME <- sd(analysis$z_ME)
mean_z_noME <- mean(analysis$z_noME)
sd_z_noME <- sd(analysis$z_noME)

results <- data.frame(mean_intercept_ME,sd_intercept_ME,mean_intercept_noME,sd_intercept_noME,mean_x1_ME,sd_x1_ME,mean_x1_noME,sd_x1_noME,mean_x2_ME,sd_x2_ME,mean_x2_noME,sd_x2_noME,mean_z_ME,sd_z_ME,mean_z_noME,sd_z_noME)

results
# Create histograms of estimations across replications
library(tidyverse)
library(patchwork)
p1 <- ggplot(analysis, aes(x=intercept_ME)) + geom_histogram() + geom_vline(aes(xintercept=2,color='red')) + theme(legend.position = "none") + ggtitle('Intercept w/ Measurement Error')
p2 <- ggplot(analysis, aes(x=intercept_noME)) + geom_histogram() + geom_vline(aes(xintercept=2,color='red')) + theme(legend.position = "none") + ggtitle('Intercept w/o Measurement Error')
inter_dist <- p1+p2

p3 <- ggplot(analysis, aes(x=x1_ME)) + geom_histogram() + geom_vline(aes(xintercept=1,color='red')) + theme(legend.position = "none") + ggtitle('x_1 w/ Measurement Error')
p4 <- ggplot(analysis, aes(x=x1_noME)) + geom_histogram() + geom_vline(aes(xintercept=1,color='red')) + theme(legend.position = "none") + ggtitle('x_1 w/o Measurement Error')
x1_dist <- p3+p4

p5 <- ggplot(analysis, aes(x=x2_ME)) + geom_histogram() + geom_vline(aes(xintercept=1,color='red')) + theme(legend.position = "none") + ggtitle('x_2 w/ Measurement Error')
p6 <- ggplot(analysis, aes(x=x2_noME)) + geom_histogram() + geom_vline(aes(xintercept=1,color='red')) + theme(legend.position = "none") + ggtitle('x_2 w/o Measurement Error')
x2_dist <- p5+p6

p7 <- ggplot(analysis, aes(x=z_ME)) + geom_histogram() + geom_vline(aes(xintercept=-1,color='red')) + theme(legend.position = "none") + ggtitle('z w/ Measurement Error')
p8 <- ggplot(analysis, aes(x=z_noME)) + geom_histogram() + geom_vline(aes(xintercept=-1,color='red')) + theme(legend.position = "none") + ggtitle('z w/o Measurement Error')
p8
z_dist <- p7+p8

inter_dist / x1_dist / x2_dist / z_dist
```
````

**Summary statistics of estimates of regression coefficients (across replications)**

| | | Intercept | X_1 | X_2 | Z |
|---|---|---|---|---|---|
| Mean | ME | 2.004 | 1.002 | 1.000 | -1.004 |
| | No ME | 1.585 | 0.785 | 0.651 | -0.170 |
| Standard deviation | ME | 0.057 | 0.047 | 0.050 | 0.101 |
| | No ME | 0.065 | 0.047 | 0.048 | 0.108 |

# Histogram plots of the distribution of estimated regression coefficients