Week 11

AUTHOR

Daniel Poisel

Task 1 - Bayesian Inference for a COVID-19 Test

Let θ be the unknown state variable:

$$\theta = \begin{cases} 1 & \text{Max has COVID-19,} \\ 0 & \text{Max does not have COVID-19.} \end{cases}$$

The doctor's prior belief based on background knowledge H is:

$$P(\theta = 1 \mid H) = 0.7, \qquad P(\theta = 0 \mid H) = 0.3.$$

The PCR test result is denoted by:

$$X = \begin{cases} 1 & \text{positive test,} \\ 0 & \text{negative test.} \end{cases}$$

The test characteristics are:

$$P(X = 1 \mid \theta = 1) = 0.95,$$

$$P(X = 1 \mid \theta = 0) = 0.4.$$

(a) Probability that Max has COVID-19 given a positive test

Using Bayes' theorem:

$$P(\theta = 1 \mid X = 1) = \frac{P(X = 1 \mid \theta = 1)\, P(\theta = 1)}{P(X = 1)}.$$

Compute the marginal probability of a positive test:

$$P(X = 1) = P(X = 1 \mid \theta = 1)\, P(\theta = 1) + P(X = 1 \mid \theta = 0)\, P(\theta = 0).$$

Insert the values:

$$P(X = 1) = 0.95 \cdot 0.7 + 0.4 \cdot 0.3 = 0.665 + 0.12 = 0.785.$$

Now compute the posterior probability:

$$P(\theta = 1 \mid X = 1) = \frac{0.95 \cdot 0.7}{0.785} \approx 0.847.$$

(b) Probability that Max does not have COVID-19 given a positive test

Using the complement rule:

$$P(\theta = 0 \mid X = 1) = 1 - P(\theta = 1 \mid X = 1).$$

$$P(\theta = 0 \mid X = 1) = 1 - 0.847 = 0.153.$$

(c) Has the probability that Max has COVID-19 increased?

The prior probability was:

$$P(\theta = 1 \mid H) = 0.7.$$

The posterior probability after observing a positive test is:

$$P(\theta = 1 \mid X = 1) \approx 0.847.$$

Since

$$P(\theta = 1 \mid X = 1) > P(\theta = 1 \mid H),$$

the probability that Max has COVID-19 has **increased**.

This happens because a positive test result is much more likely when Max has COVID-19 than when he does not, so the observation provides evidence in favor of the disease.

Task 2

From Question 1, the posterior distribution after the first positive test $X_1 = 1$ is:

$$P(\theta = 1 \mid X_1 = 1) \approx 0.847, \qquad P(\theta = 0 \mid X_1 = 1) \approx 0.153.$$

The second test X_2 has the following characteristics:

$$P(X_2 = 1 \mid \theta = 1) = 0.99,$$

$$P(X_2 = 1 \mid \theta = 0) = 0.04.$$

(a)

We compute the conditional probability of a positive second test using the law of total probability:

$$P(X_2 = 1 \mid X_1 = 1) = P(X_2 = 1 \mid \theta = 1)P(\theta = 1 \mid X_1 = 1) + P(X_2 = 1 \mid \theta = 0)P(\theta = 0 \mid X_1 = 1).$$

Insert the numerical values:

$$P(X_2 = 1 \mid X_1 = 1) = 0.99 \cdot 0.847 + 0.04 \cdot 0.153.$$

$$P(X_2 = 1 \mid X_1 = 1) = 0.8385 + 0.0061 = 0.8446.$$

Thus, the probability of a negative second test result is:

$$P(X_2 = 0 \mid X_1 = 1) = 1 - P(X_2 = 1 \mid X_1 = 1) = 1 - 0.8446 = 0.1554.$$

So the conditional distribution of X_2 given $X_1 = 1$ is:

$$X_2 \mid X_1 = 1 \sim \begin{cases} 1 & \text{with probability } 0.8446, \\ 0 & \text{with probability } 0.1554. \end{cases}$$

(b)

We now update our belief about θ using Bayes' theorem:

$$P(\theta = 1 \mid X_1 = 1, X_2 = 0) = \frac{P(X_2 = 0 \mid \theta = 1)\,P(\theta = 1 \mid X_1 = 1)}{P(X_2 = 0 \mid X_1 = 1)}.$$

First compute the likelihoods:

$$P(X_2 = 0 \mid \theta = 1) = 1 - 0.99 = 0.01,$$

$$P(X_2 = 0 \mid \theta = 0) = 1 - 0.04 = 0.96.$$

Next compute the marginal probability:

$$P(X_2 = 0 \mid X_1 = 1) = 0.01 \cdot 0.847 + 0.96 \cdot 0.153.$$

$$P(X_2 = 0 \mid X_1 = 1) = 0.00847 + 0.14688 = 0.15535.$$

Now compute the posterior probability:

$$P(\theta = 1 \mid X_1 = 1, X_2 = 0) = \frac{0.01 \cdot 0.847}{0.15535} \approx 0.055.$$

Using the complement rule:

$$P(\theta = 0 \mid X_1 = 1, X_2 = 0) = 1 - 0.055 = 0.945.$$

After a positive first test and a negative second test, the posterior distribution of θ is:

$$\theta \mid (X_1 = 1, X_2 = 0) \sim \begin{cases} 1 & \text{with probability } 0.055, \\ 0 & \text{with probability } 0.945. \end{cases}$$

Despite the strong evidence from the first test, the highly accurate negative second test substantially reduces the probability that Max has COVID-19.

Task 3

Bob has three possible modes of transport to work:

$$T \in \{1, 2, 3\},$$

where:

- $T = 1$: car
- $T = 2$: bus
- $T = 3$: train

The probability that Bob is late depends on the mode of transport:

$$P(L \mid T = 1) = 0.5,$$

$$P(L \mid T = 2) = 0.2,$$

$$P(L \mid T = 3) = 0.01,$$

where L denotes the event that Bob is late.

The line manager assigns equal prior probabilities to each transport mode:

$$P(T = 1) = P(T = 2) = P(T = 3) = \frac{1}{3}.$$

Using the law of total probability:

$$P(L) = \sum_{t=1}^{3} P(L \mid T = t) \, P(T = t).$$

Substituting the values:

$$P(L) = \frac{1}{3}(0.5 + 0.2 + 0.01) = \frac{0.71}{3} \approx 0.2367.$$

Using Bayes' theorem:

$$P(T = 1 \mid L) = \frac{P(L \mid T = 1) \, P(T = 1)}{P(L)}.$$

Substitute the values:

$$P(T = 1 \mid L) = \frac{0.5 \cdot \frac{1}{3}}{0.2367}.$$

$$P(T = 1 \mid L) \approx \frac{0.1667}{0.2367} \approx 0.704.$$

The manager's estimate of the probability that Bob came to work by car is:

$$P(T = 1 \mid L) \approx 0.704.$$

This means that, given Bob was late, there is approximately a **70.4% probability** that he traveled by car.

Task 4

a)

i)

```
set.seed(123)   # for reproducibility

# Parameters

mu <- 5
sigma <- 1      # known standard deviation
n <- 50         # sample size per replication
B <- 100        # number of replications
```

```
  z <- 1.96        # z-score for 95% CI

  # Storage for CI bounds

  lower <- numeric(B)
  upper <- numeric(B)

  # Simulation loop

  for (b in 1:B) {
    x <- rnorm(n, mean = mu, sd = sigma)      # generate 50 points
    xbar <- mean(x)                           # sample mean
    lower[b] <- xbar - z * sigma / sqrt(n)    # lower bound of CI
    upper[b] <- xbar + z * sigma / sqrt(n)    # upper bound of CI
  }

  # Combine results

  ci <- cbind(mean(lower), mean(upper))
  ci
```

```
         [,1]      [,2]
[1,] 4.722245 5.276616
```

ii)

```
  # Plot setup

  plot(
    NULL,
    xlim = c(min(lower), max(upper)),   # x-axis spans all CIs
    ylim = c(1, B),                      # y-axis for 100 replications
    xlab = expression(mu),
    ylab = "Replication",
    main = "95% Confidence Intervals for μ"
  )

  # Draw horizontal line at true μ

  abline(v = mu, lwd = 2, col = "blue")

  # Loop to plot each CI

  for (i in 1:B) {
  col_i <- if (lower[i] <= mu & upper[i] >= mu) "black" else "red"
  segments(lower[i], i, upper[i], i, col = col_i, lwd = 2)
  }
```
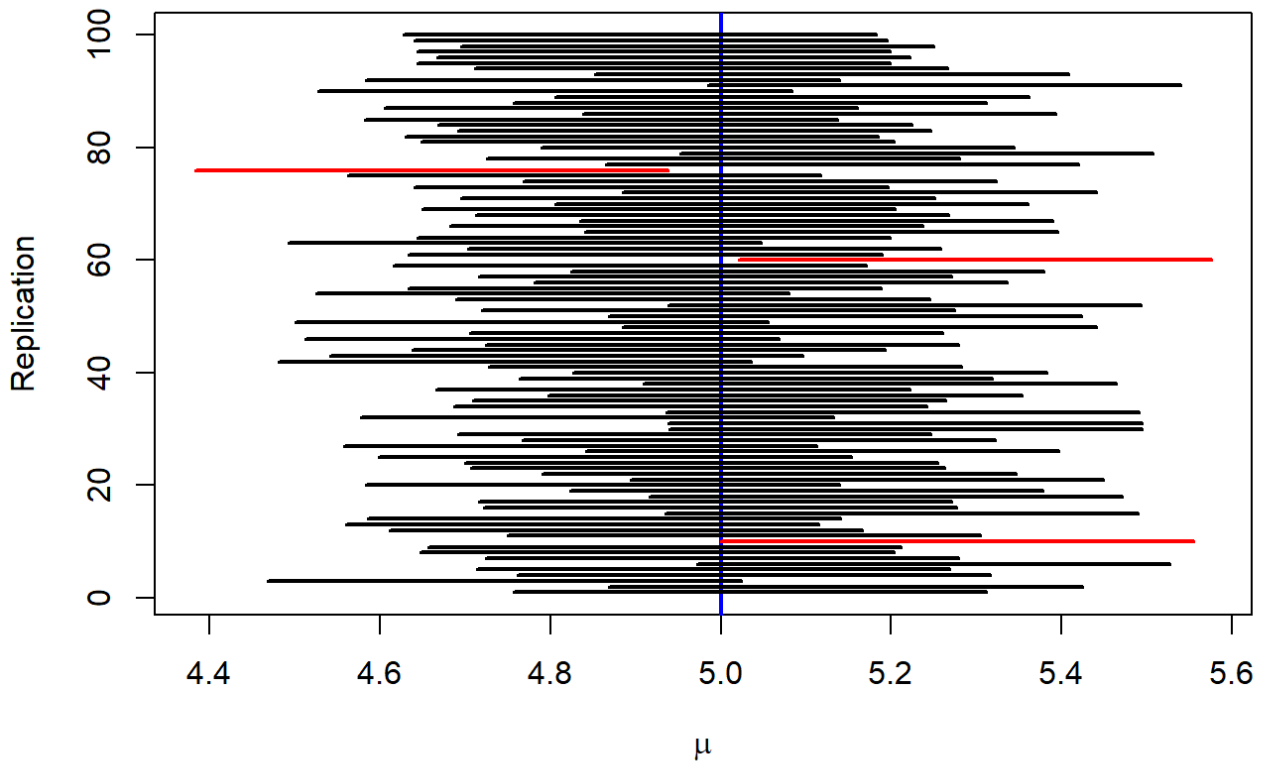
95% Confidence Intervals for μ

b)

ii)

"If you conduct your experiment 100 times and construct 95% confidence interval in each of them, it is expected that 95 confidence intervals out of these 100 would contain the true value".