BINF-6310-stats-lab3

Gloria Jackson

2026-02-05

Problem 1

Problem 1: Question 1

You walk into the "occasionally dishonest casino" with prior probabilities and likelihoods set to the values in slides 21-25 of lecture #5. (that is, a "loaded" die has a 10% chance of getting a 1-5 and a 50% chance of getting a 6 but 99% of the dice are fair)

You pick up one die and with it roll:

2 3 2 6 3 5 6 2 6 6 2 6 6 2 3 6 6 6 5 6 6 5 6 6 6 6 6 4 6 3 3 3 6 6 5 6 6

Make a graph of the posterior probability that you have picked up a loaded die as a function of the number of times you have rolled the die.

Show your code. . .

You can represent the rolls as data<-c(2,3,2,6,3,5,6,2,6,6,2,6,6,2,3,6,6,6,5,6,6,5,6,6,6,6,6,4,6,3,3,3,6,6,5,6,6)

```r
rm(list=ls())

# priors
p_loaded <- 0.01
p_fair <- 0.99

# observed rolls
data<-c(2,3,2,6,3,5,6,2,6,6,2,6,6,2,3,6,6,6,5,6,6,5,6,6,6,6,6,4,6,3,3,3,6,6,5,6,6)

# number of rolls
n <- length(data)

# initialize likelihoods (start at 1)
lik_loaded <- 1
lik_fair <- 1

# store posterior after each roll
posterior <- numeric(n)

# iterate thru rolls
for (i in 1:n) {
  roll <- data[i]
  # update likelihoods based on roll
  if (roll == 6) {
```
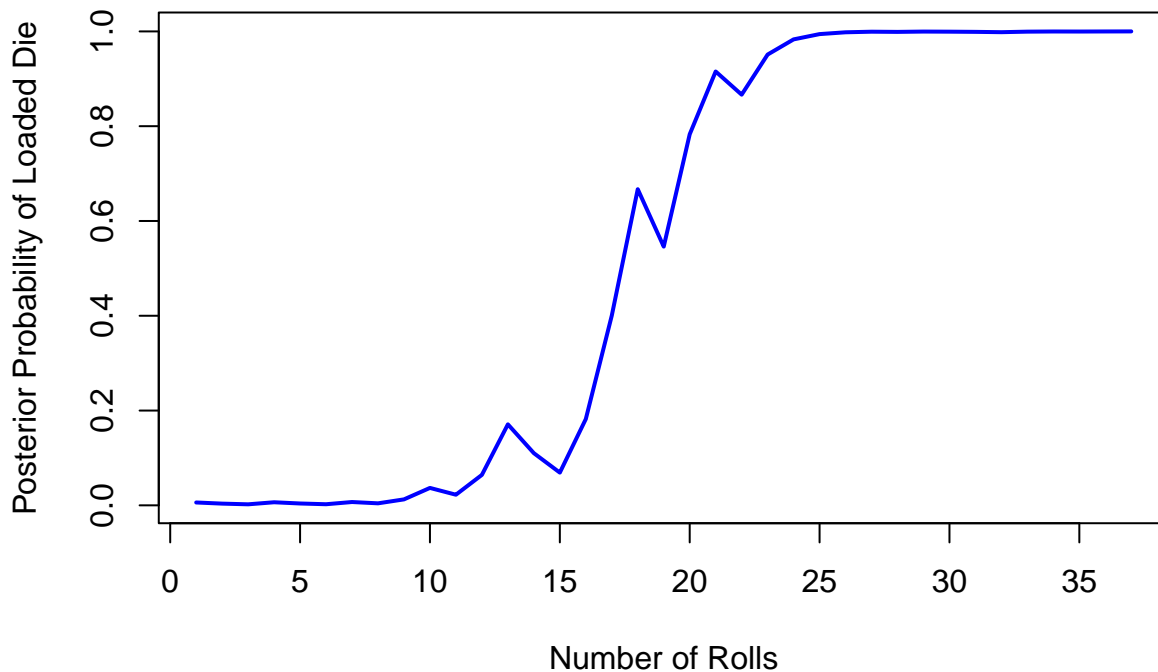
```
    lik_loaded <- lik_loaded * 0.5
    lik_fair <- lik_fair * (1/6)
  } else {
    lik_loaded <- lik_loaded * 0.1
    lik_fair <- lik_fair * (1/6)
  }
  # posterior
  posterior[i] <- (lik_loaded * p_loaded) / ((lik_loaded * p_loaded)
                                              + (lik_fair * p_fair))
}

# plot
plot(posterior, type="l",lwd=2,col="blue",xlab="Number of Rolls",
     ylab="Posterior Probability of Loaded Die",
     main="Posterior Probability That Die is Loaded")
```

Posterior Probability That Die is Loaded

Problem 1: Question 2

How many times on average would you need to roll a loaded die to be 99.999% sure that it was loaded?
(Show your work)

```
rm(list=ls())
set.seed(1) # for reproducibility

# number of simulated experiments
n_sims <- 1000
# store how many rolls are needed in each experiment
roll_counts <- numeric(n_sims)
```

```r
# loop over simulated experiments
for (i in 1:n_sims) {
  lik_loaded <- 1
  lik_fair <- 1
  posterior <- 0
  rolls <- 0

  while (posterior < 0.99999) {
    # simulate 1 roll from a loaded die
    roll <- sample(c(1,2,3,4,5,6), 1, prob = c(0.1,0.1,0.1,0.1,0.1,0.5))
    # count this roll
    rolls <- rolls + 1
    # update likelihoods based on observed roll
    if (roll == 6) {
    lik_loaded <- lik_loaded * 0.5
    lik_fair <- lik_fair * (1/6)
  } else {
    lik_loaded <- lik_loaded * 0.1
    lik_fair <- lik_fair * (1/6)
  }
    # Bayes' rule to update posterior probability
    posterior <- (lik_loaded * 0.01) / ((lik_loaded * 0.01) + (lik_fair*0.99))
  }
  # record how many rolls were needed in this experiment
  roll_counts[i] <- rolls
}

# average number of rolls required to reach 99.999% certainty
mean(roll_counts)
```

```
## [1] 56.745
```

Problem 2

You are consulting for a hospital. They have a diagnostic test for a disease with a known background prevalence of 0.1%.

The test has the following properties: p(positive result | person has disease) = 0.91 p(negative result| person does not have disease) = 0.84

The cost of running the test one time is $1. The test can be repeated for each patient and the results of the test are independent of one another allowing for Bayesian updates. The test always yields a positive or negative result.

The requirement of the hospital is that the test is repeated for each patient until a Bayesian posterior of at least 0.99999 is reached.

Problem 2: Question 1

Run simulations for a patient with the disease. About how many times on average must the test be repeated to achieve the hospital's requirements?

```r
rm(list=ls())
set.seed(1) # reproducibility insurance

n_sims <- 1000 # number of simulations
# store number of tests needed per patient
test_counts <- numeric(n_sims)

# loop over simulated patients WITH disease
for (i in 1:n_sims) {
  # initializing likelihoods & counters
  lik_disease <- 1
  lik_no_disease <- 1
  posterior <- 0.001
  tests <- 0

  while (posterior < 0.99999) {
    # simulate 1 test result for a patient WITH disease
    test_result <- sample(c("pos", "neg"), 1, prob = c(0.91, 0.09))
    # count this test
    tests <- tests + 1

    # update likelihoods based on test outcome
    if (test_result == "pos") {
      lik_disease <- lik_disease * 0.91
      lik_no_disease <- lik_no_disease * 0.16
    } else {
      lik_disease <- lik_disease * 0.09
      lik_no_disease <- lik_no_disease * 0.84
    }
    # Bayes' rule to update posterior probabilitiy
    posterior<-(lik_disease*0.001)/((lik_disease*0.001)+(lik_no_disease*0.999))
  }
  # record how many tests were required
  test_counts[i] <- tests
}

# average number of required tests
mean(test_counts)
```

```
## [1] 13.96
```

Problem 2: Question 2

Repeat the simulations for a patient without the disease. About how many times on average must the test be repeated to achieve the hospital's requirements?

```r
rm(list=ls())
set.seed(1) # for reproducibility

n_sims <- 1000 # number of simulations
# store the number of tests needed per patient
test_counts <- numeric(n_sims)
```

```r
# loop over simulated patients WITHOUT disease
for (i in 1:n_sims) {
  # initializing likelihoods and counters using log
  lik_disease <- 1
  lik_no_disease <- 1
  posterior <- 0.001 # start at prior (?)
  tests <- 0

  while (posterior > 0.00001 && posterior < 0.99999) {
    # simulate 1 test result for patient WITHOUT disease
    test_result <- sample(c("pos","neg"),1,prob = c(0.16,0.84))
    #count this test
    tests <- tests + 1

    # update likelihoods based on test outcome
    if (test_result == "pos") {
      lik_disease <- lik_disease * 0.91
      lik_no_disease <- lik_no_disease * 0.16
    } else {
      lik_disease <- lik_disease * 0.09
      lik_no_disease <- lik_no_disease * 0.84
    }

    # Bayes' rule to update posterior probability
    posterior<-(lik_disease*0.001)/((lik_disease*0.001)+(lik_no_disease*0.999))
  }
  # record required number of tests
  test_counts[i] <- tests
}

# average number of required tests
mean(test_counts)
```

```
## [1] 3.87
```

Problem 2: Question 3

The hospital plans to run the test on one million patients per year. At a cost of $1 per test, about how much should the hospital budget to run these tests? (That is to say, for a million patients, how many tests can the hospital anticipate running?)

Show your work/code/justification for all answers.

```r
# avg number of tests for patient WITH disease
mu_D <- 13.96 # result from problem 2: question 1
# avg number of tests for patient WITHOUT disease
mu_N <- 3.87 # result from problem 2: question 2

expected_tests_per_patient <- 0.001 * mu_D + 0.999 * mu_N

total_tests <- 1000000 * expected_tests_per_patient
```

```r
total_budget <- total_tests * 1
total_budget
```

```
## [1] 3880090
```

Disclaimer: AI was used to help generate some code.