Code

Lab 3 | BINF 6310 | Spring 2020 | Jon Lee

Part 1:

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 = rate of true positive result p(negative result| person does not have disease) = 0.84 = rate of true negative result

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 0.9999 is reached for >=95% of patients.

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?

Hide

```
#get data function
getData <- function(likelihood, numPts)
{
  d <- vector(mode = "integer", length = numPts)

  for(i in 1:numPts)
  {
    if(runif(1) <= likelihood[1])
    {
      d[i] <- 1
    }
    else
    {
      d[i] <- 2
    }
  }

  return(d)
}

#estmated power function
estimatedPower <- function(prob, likelihood_1, likelihood_2, numTests, numSimulationsPer
Cycle)
{
  t <- 1:numTests
  s <- numSimulationsPerCycle

  avgPosterior <- vector(length = length(t))
  estPower <- vector(length = length(t))

  for(i in t)
  {
    posteriorValues <- vector(length = s)

    for(j in 1:s)
    {
      p <- prob

      data <- getData(likelihood_1, t[i])

      for(k in 1:length(data))
      {
        denom <- likelihood_2[data[k]]*p[2]+likelihood_1[data[k]]*p[1]

        p[1] = p[1]*likelihood_1[data[k]]/denom
        p[2] = p[2]*likelihood_2[data[k]]/denom
      }

      posteriorValues[j] = p[1]
    }

    avgPosterior[i] = mean(posteriorValues)
    estPower[i] <- sum(posteriorValues >= 1-0.0001)/s
```

```
  }

  return(estPower)
}
```

Hide

```
#probability of having the disease and being healthy
probOfDisease <- c(0.001, 0.999)

#likelibhoods given a positive, negative result
likelihoodGivenDisease <- c(0.91, 0.09)
likelihoodGivenHealthy <- c(0.16, 0.84)

diseased <- estimatedPower(probOfDisease, likelihoodGivenDisease, likelihoodGivenHealth
y, 25, 10000)
print(diseased)
```

```
 [1] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.3943
[11] 0.3535 0.7075 0.6748 0.8678 0.8499 0.8273 0.9394 0.9310 0.9761 0.9682
[21] 0.9919 0.9882 0.9979 0.9962 0.9945
```

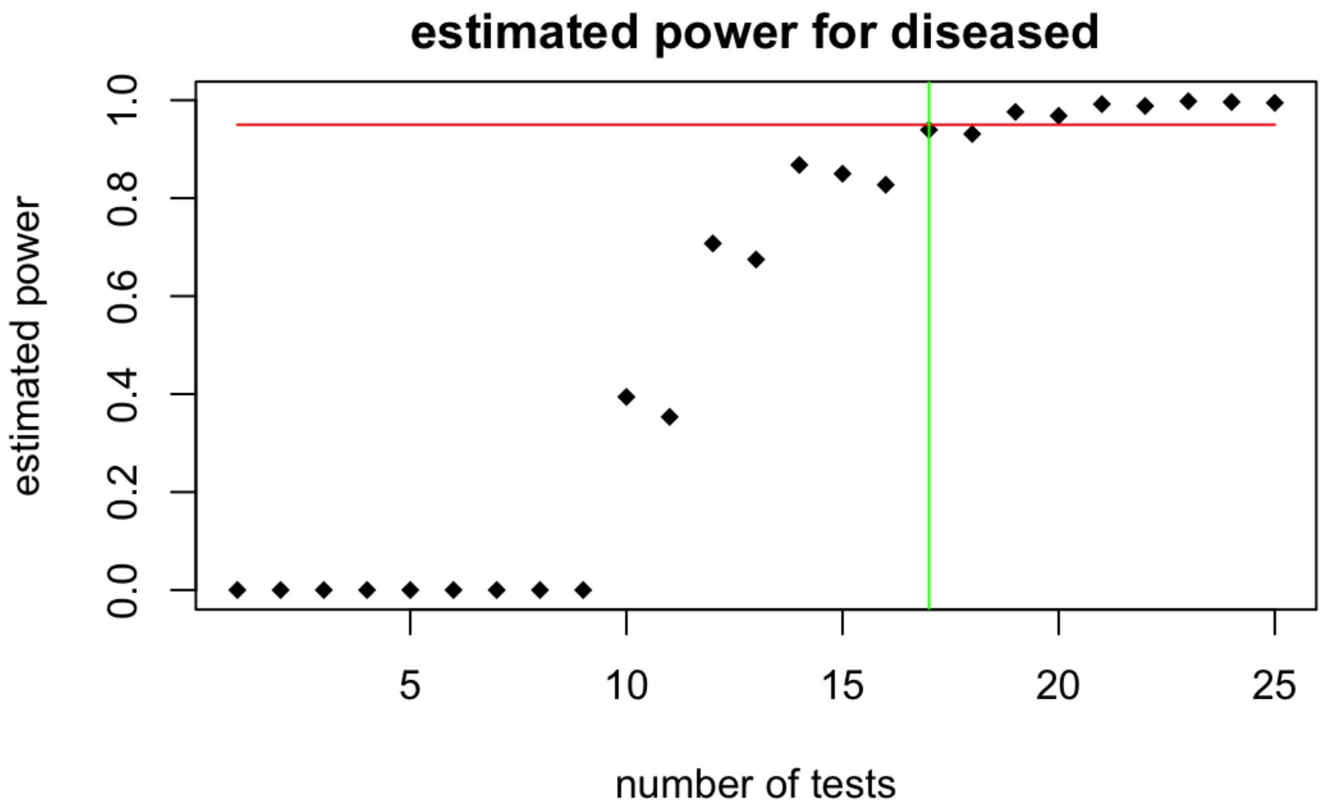
Hide

```
numberOfTests <-1:25

plot(numberOfTests, diseased, main = "estimated power for diseased", xlab = "number of t
ests", ylab = "estimated power", pch = 18)
lines(numberOfTests, rep(0.95,length(numberOfTests)), col = "red")
```

Hide

```
abline(v = 17, col = "green")
```



Answer:

It takes approximately 17 tests to reach the hospital requirement that ≥95% of pateints have a Bayesian posterior probability of ≥ 0.9999, for patients with the disease testing positive.

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?

Hide

```
#probability of having the disease and being healthy
probOfHealthy <- c(0.999, 0.001)

#need to flip the probability vector here since we are asking about healthy patients now

#likelibhoods given a positive, negative result
likelihoodGivenDisease <- c(0.91, 0.09)
likelihoodGivenHealthy <- c(0.16, 0.84)

#reverse the order of healthy and diseased likelihoods since we are asking about healthy
patients

healthy <- estimatedPower(probOfHealthy, likelihoodGivenHealthy, likelihoodGivenDisease,
25, 10000)
print(healthy)
```

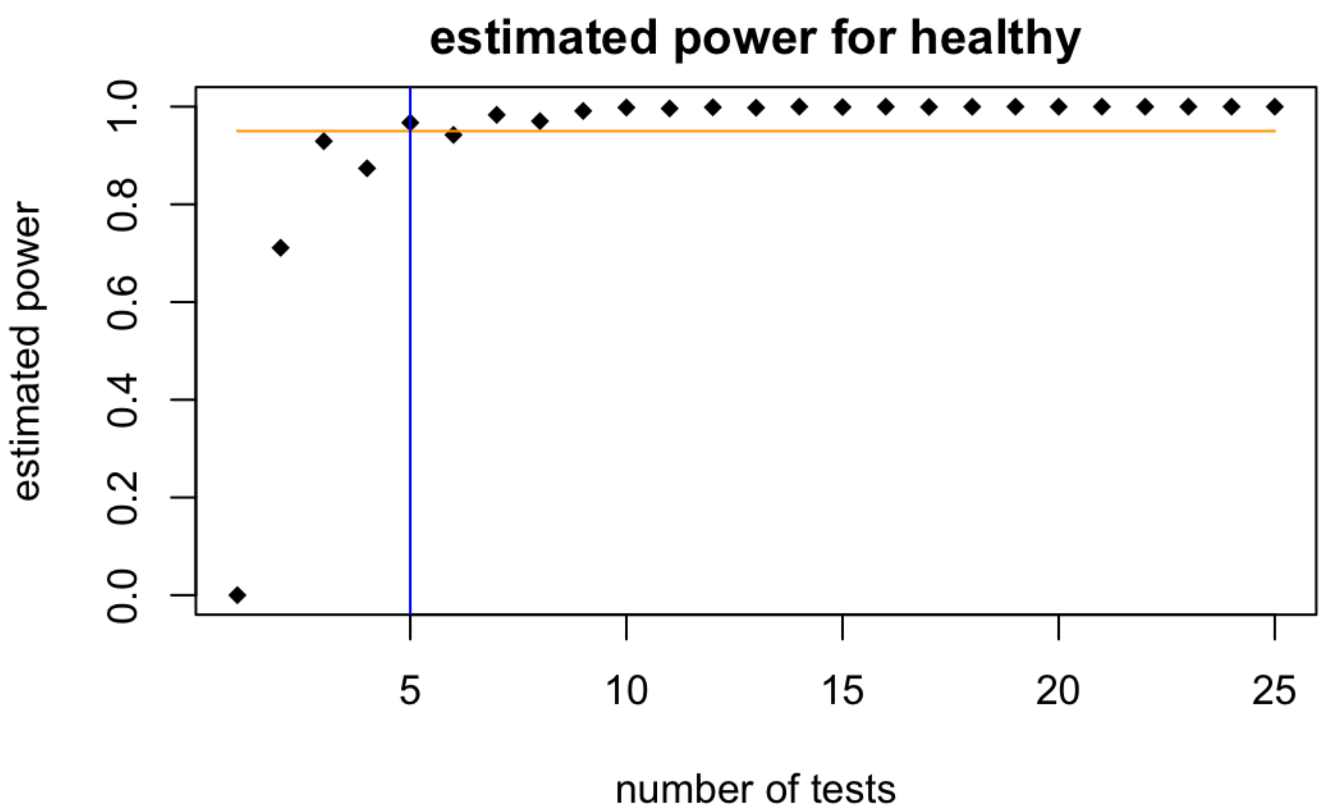
```
 [1] 0.0000 0.7111 0.9293 0.8738 0.9672 0.9422 0.9832 0.9703 0.9912 0.9982
[11] 0.9963 0.9987 0.9978 1.0000 0.9991 1.0000 0.9994 0.9997 1.0000 0.9999
[21] 1.0000 0.9999 1.0000 1.0000 0.9999
```

Hide

```
plot(numberOfTests,  healthy, main = "estimated power for healthy", xlab = "number of te
sts", ylab = "estimated power", pch = 18)
lines(numberOfTests, rep(0.95, length(numberOfTests)), col = "orange")
```

Hide

```
abline(v = 5, col = "blue")
```



Answer:

It takes approximately 5 tests to reach the hospital requirement that ≥95% of pateints have a Bayesian posterior probability of ≥ 0.9999, for healthy patients that test negative.

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?)

Hide

```
costOfTest = 1
numberOfPatients = 1000000
numberOfTestsPerPatient = 17

projectedCost = costOfTest*numberOfPatients*numberOfTestsPerPatient
print(projectedCost)
```

```
[1] 1.7e+07
```

Answer:

The hospital should budget 17 million dollars per year if they plan to run the test of 1 million patients. The hospital should budget/test for the condition that the paients are diseases and test positive rather than the opposing condition, becauase of the prevalence of the disease is so small. Although this is more costly, it is a better identifier because individuals will not be tested unless they fall under the category that they could have the disease.

Part 2:

Another manufacturer approaches the hospital with an improved, but more expensive, test with the following properties

p(positive result | person has disease) = 0.96 p(negative result| person does not have disease) = 0.95

1. With this test, how many times on average must the test be repeated to achieve the hospital's requirements for patients with and without the disease?

Hide

```
#likelibhoods given a positive, negative result for the new test
likelihoodGivenDisease_newTest <- c(0.96, 0.04)
likelihoodGivenHealthy_newTest <- c(0.05, 0.95)

#probabilty of disease is still the same

diseased_newTest <- estimatedPower(probOfDisease, likelihoodGivenDisease_newTest, likeli
hoodGivenHealthy_newTest, 25, 10000)
print(diseased_newTest)
```

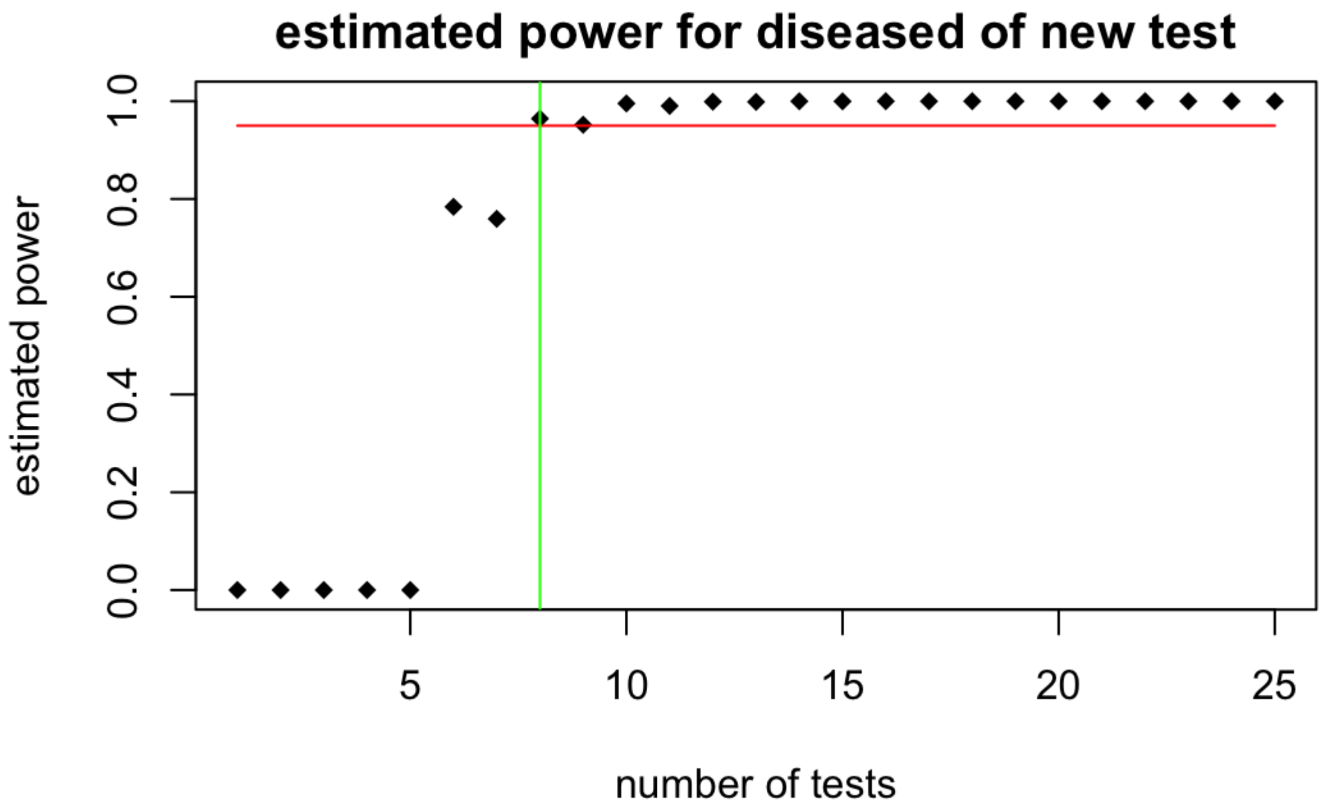
```
 [1] 0.0000 0.0000 0.0000 0.0000 0.0000 0.7840 0.7594 0.9644 0.9519 0.9954
[11] 0.9904 0.9990 0.9986 0.9999 0.9997 1.0000 0.9999 1.0000 1.0000 1.0000
[21] 1.0000 1.0000 1.0000 1.0000 1.0000
```

Hide

```
plot(numberOfTests, diseased_newTest, main = "estimated power for diseased of new test",
xlab = "number of tests", ylab = "estimated power", pch = 18)
lines(numberOfTests, rep(0.95, length(numberOfTests)), col = "red")
```

Hide

```
abline(v = 8, col = "green")
```



Answer:

It takes approximately 8 tests to reach the hospital requirement that ≥95% of pateints have a Bayesian posterior probability of ≥ 0.9999, for patients with the disease testing positive with the new test.

Hide

```
#use the reverse healthy probability vector here since we are asking about healthy patie
nts now and reverse the order of healthy and diseased likelihoods since we are asking ab
out healthy patients for the new test

healthy_newTest <- estimatedPower(probOfHealthy, likelihoodGivenHealthy_newTest, likelih
oodGivenDisease_newTest, 25, 10000)
print(healthy_newTest)
```

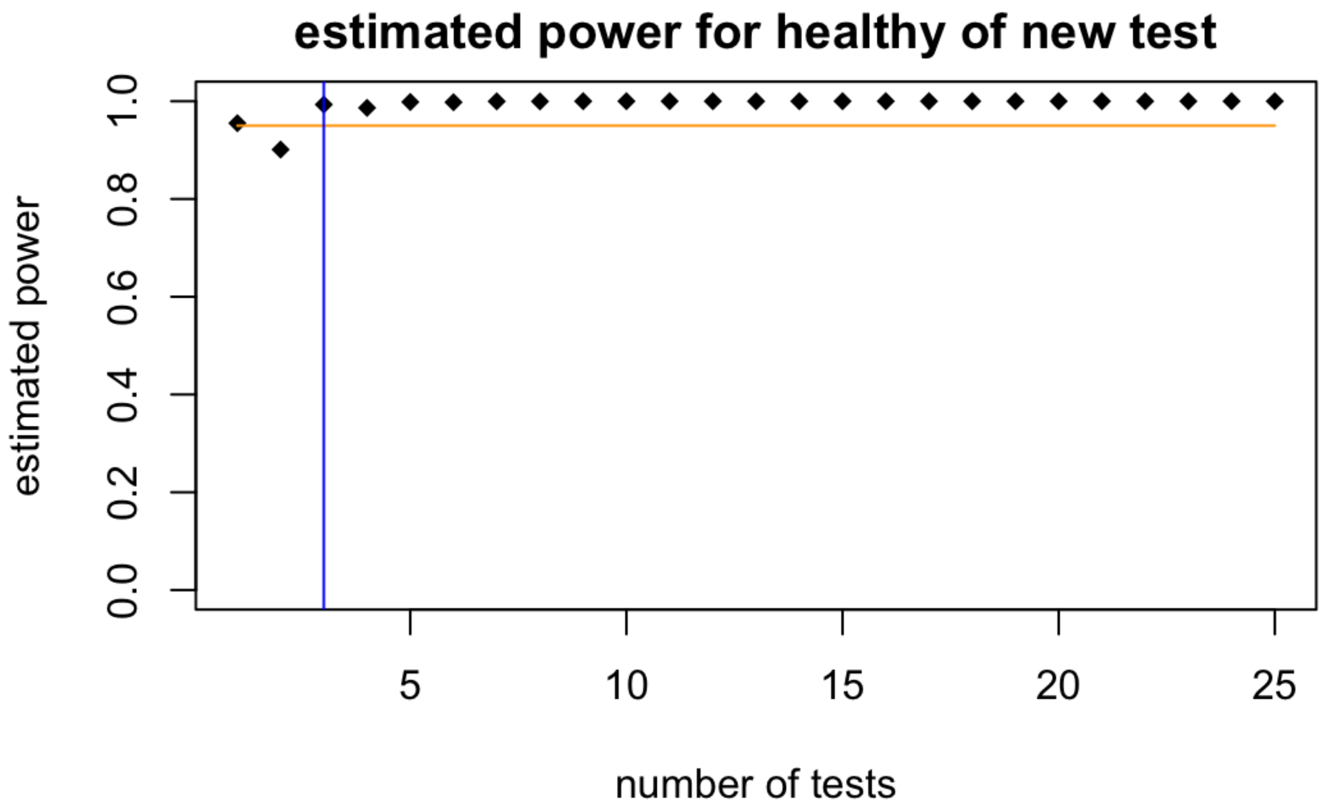
```
 [1] 0.9550 0.9011 0.9932 0.9864 0.9984 0.9979 0.9997 0.9996 1.0000 1.0000
[11] 1.0000 1.0000 0.9999 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
[21] 1.0000 1.0000 1.0000 1.0000 1.0000
```

Hide

```
plot(numberOfTests, healthy_newTest, main = "estimated power for healthy of new test", x
lab = "number of tests", ylab = "estimated power", pch = 18, ylim = c(0,1))
lines(numberOfTests, rep(0.95, length(numberOfTests)), col = "orange")
```

Hide

```
abline(v = 3, col = "blue")
```



Answer:

It takes approximately 3 tests to reach the hospital requirement that ≥95% of pateints have a Bayesian posterior probability of ≥ 0.9999, for healthy patients that test negative.

2. Considering only the cost of the test, and assuming the hospital will screen one million patients with a background prevalence of 0.1%, at about what price point for running the test one time will the hospital save money by switching to the new test?

Hide

```
#new test parametrics
numberOfTestsPerPatient_newTest = 8

#what the hospital should pay for the new test per capita, assuming they don't pay more
 than 17 million in total

costOfTest_newTest = projectedCost/(numberOfTestsPerPatient_newTest*numberOfPatients)
print(costOfTest_newTest)
```

```
[1] 2.125
```

Answer:

The hospital should not pay more than rouhgly $2.12 for the new test.