

Problem Set 1

Michael Fryer

2025-11-07

1. COVID Home Test

In a clinical trial, the BinaxNOW home COVID-19 antigen test correctly gave a positive result 75.5% of the time and correctly gave a negative result 99.5% of the time. For the next set of questions, assume that presence of the antigen suffices for a person to have COVID-19 and absence of the antigen suffices for that person to not have COVID-19. Finally, assume that 10% of the people in your community are currently infected with COVID-19. (This is your base rate for exposure.)

```
# Data from above
p_infected <- 0.1
p_not_infected <- 1 - p_infected

p_pos_g_infected <- 0.755
p_neg_g_infected <- 1 - p_pos_g_infected
p_neg_g_not_infected <- 0.995
p_pos_g_not_infected <- 1 - p_neg_g_not_infected
```

a) Suppose you take a BinaxNOW COVID-19 antigen test. What is the probability of an administered BinaxNOW test returning to you a positive result?

The probability of an administered BinaxNOW test returning to you a positive result is equal to $P(T = 1) = P(T = 1|C = 1)P(I = 1) + P(T = 1|C = 0)P(C = 0)$ where T refers to the test result and C refers to the presence of the antigen.

```
p_pos <- p_pos_g_infected * p_infected + p_pos_g_not_infected * p_not_infected
cat("Probability of getting a positive result: ", p_pos)
```

```
## Probability of getting a positive result: 0.08
```

b) Suppose the infection rate in New Zealand is 1.5% and a New Zealander takes a BinaxNOW test. What is the probability that this test will return a positive result?

This is the same as above, however we use different values for our priors.

```
p_infected_nz <- 0.015
p_not_infected_nz <- 1 - p_infected_nz

p_pos <- (
  p_pos_g_infected * p_infected_nz + p_pos_g_not_infected * p_not_infected_nz
)
cat("Probability of getting a positive result (in New Zealand): ", p_pos)
```

```
## Probability of getting a positive result (in New Zealand): 0.01625
```

c) A competitor offers a test with a sensitivity 90% but specificity 99.0% (vs BinaxNOW's 99.5%).

- Calculate $P(C = 1|T = 1)$ for both tests.

- The competitor's test costs 2x more. For which base rates (if any) would you prefer the competitor's test?
- Plot $P(C = 1|T = 1)$ vs base rate for both tests on the same graph.
- Write 2-3 sentences explaining at which base rates each test is preferable and why the preference changes (or doesn't change).

d) Suppose the competitor offers a second generation test to you with a sensitivity 82% and specificity 99.7%.

- Calculate $P(C_1|T_1)$ for BinaxNOW vs the competitor's generation 2 test.
- The competitor's new test also costs 2x more than BinaxNOW. For which base rates (if any) would you prefer the competitor's new test?
- Plot $P(C_1|T_1)$ vs base rate for both tests on the same graph.
- Write 2-3 sentences explaining at which base rates each test is preferable and why the preference changes (or doesn't change).

2. Computing Probabilities

Implement and run the following chunk of code to create a distribution, samples. Create a plot of that distribution. Then, where called for, write a short line of R code to compute an answer to each question. Analytical solutions will not be accepted.

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prior <- rep(1, 1000)
likelihood <- dbinom( 6 , size=10 , prob=p_grid )
posterior <- likelihood * prior
posterior <- posterior / sum(posterior)
set.seed(712)
samples <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

- How much posterior probability lies below $p = 0.5$?
- How much posterior probability lies above $p = 0.8$?
- How much posterior probability lies between $p = 0.2$ and $p = 0.8$?
- 20% of the posterior probability lies below which value of p ?

3. Swing Voters

Write your own R code chunks to answer the following questions. Analytical solutions will not suffice.

a) Imagine a country where there are only two political parties, Red and Blue, which divide the electorate equally. One difference between registered Blue voters and registered Red voters is their willingness to vote for the opposing party's candidate. Blue voters vote Red 20% of the time, otherwise they vote Blue. Red voters vote Blue 10% of the time, otherwise they vote Red. Voters who switch are called swing voters.

Smith was a swing voter in the last election but you do not know whether he is Red or Blue. (Nobody changes parties.) What is the probability that Smith will be a swing voter in the next election? Explain your reasoning. Assume the prior probability of a person being in Blue is 0.265.

b) Now imagine a country where there are three political parties: Red, Blue, and Green. Red voters vote Blue 10% of the time, vote for Green 5% of the time, and vote their own party, Red, 85% of the time. Blue voters vote Red 15% of the time, Green 5% of the time, and their own party the remaining 80% of the time. Finally, Green votes Blue 20% of the time, Red 10% of the time, and Green the remainder. The electorate is evenly among the three parties.

What is the probability that a swing voter in the last election between Red, Blue, and Green, will be a swing voter in the next election? (Like before, nobody changes parties.) Explain your reasoning.

4. Reflection

Look back at problems 1 to 3. In each case, you updated beliefs based on observations:

- In problem 1, test result → disease status.
- In problem 2, data → parameter value.
- In problem 3, past behavior → future behavior.

Write a 3-4 sentence paragraph explaining what these three problems have in common from a Bayesian perspective. What role do priors play in each?

5. AI Declaration

Please declare your collaborators in the class and how you used AI (if at all) to complete this assignment. If you used AI, include the prompts you used and explain what you learned from its responses that you didn't understand initially.