STAT641 - Homework 2

Marcel Gietzmann-Sanders

Problem 1

a. I suppose I would wonder whether this measurement is all that extreme given my initial hypothesis. More precisely I'd wonder what the probability finding at least this many hypertensive people in a selection of five would be given my hypothesis.

```
sum(dbinom(c(4, 5), 5, 0.3))
```

```
## [1] 0.03078
```

Given this is exceedingly small (only ~3% likely given my hypothesis) I would reject my initial hypothesis and decide that θ is likely higher than 0.3. I wouldn't necessarily say $\theta = 0.2$ given the small sample size but it's certainly higher than 0.3.

b. At this point, given the large sample size I would likely estimate $\theta = \frac{400}{1000} = 0.4$. I have enough data here that whatever ill conceived notions I had at the beginning had better be laid to rest.

Problem 2

Let's start by gathering what we know.

$$P(Y = 1|\theta = 1) = 0.975$$

$$P(Y = 0|\theta = 0) = 0.95$$

$$P(\theta = 1) = 0.01$$

a. What we are concerned with here is the probability that someone has the disease given they tested positive - $P(\theta = 1|Y = 1)$.

Using Bayes rule we have:

$$P(\theta = 1|Y = 1) = \frac{P(Y = 1|\theta = 1)P(\theta = 1)}{P(Y = 1)}$$

We know $P(Y = 1|\theta = 1)$ and $P(\theta = 1)$ so all that remains is to determine $P(Y = 1)$

$$P(Y = 1) = P(Y = 1|\theta = 0)P(\theta = 0) + P(Y = 1|\theta = 1)P(\theta = 1)$$

$$P(Y = 1) = [1 - P(Y = 0|\theta = 0)]\,[1 - P(\theta = 1)] + P(Y = 1|\theta = 1)P(\theta = 1)$$

$$P(Y = 1) = 0.05(0.99) + 0.975(0.01) = 0.05925$$

Therefore we have:

$$P(\theta = 1|Y = 1) = \frac{0.975(0.01)}{0.05925} \approx 0.165$$

So of those who test postive for the disease approximately 16.5% of them will actually have the disease. Therefore there is reason for concern if you test positive, but more often than not those who test positive won't have the disease.

b. In this case we're interested in $P(\theta = 0|Y = 0)$

$$P(\theta = 0|Y = 0) = \frac{P(Y = 0|\theta = 0)P(\theta = 0)}{P(Y = 0)}$$

$$P(\theta = 0|Y = 0) = \frac{P(\theta = 0|Y = 0)\,[1 - P(\theta = 1)]}{[1 - P(Y = 0)]}$$

$$P(\theta = 0|Y = 0) = \frac{0.95(0.99)}{0.94075} \approx 0.9997$$

So if someone takes the test and gets a negative there is an approximately 99.97% chance that they are indeed disease free. If you get a negative there's more or less no need to be worried.

Problem 3

Let L stand in for our color of labrador with 1, 2, 3 being yellow, chocolate, and black respectively. Let $B = 1$ mean they love to play fetch and $B = 0$ mean that they don't. We are interested in the following:

$$P(L = x|B = 1)$$

In general:

$$P(L = x|B = 1) = \frac{P(B = 1|L = x)P(L = x)}{P(B = 1)}$$

We have the numerators for our various colors so all we really need to determine is $P(B = 1)$. But this is just:

$$P(B = 1) = \sum_x P(B = 1|L = x)P(L = x)$$

$$P(B = 1) = 0.95(0.5) + 0.92(0.3) + 0.9(0.2) = 0.931$$

So now we have:

$$P(L = 1|B = 1) = \frac{0.95(0.5)}{0.931} \approx 0.51$$

$$P(L = 2|B = 1) = \frac{0.92(0.3)}{0.931} \approx 0.3$$

$$P(L = 3|B = 1) = \frac{0.9(0.2)}{0.931} \approx 0.19$$

So if we know the dog likes to play fetch there's a 51% probability it's a yellow, 30% that its a chocolate lab, and 19% that its a black lab. Clearly loving to play fetch is not particularly discriminating.

Problem 4

a. In this case the likelihood is just:

$$p(3|\theta) = \binom{8}{3}\theta^3(1 - \theta)^5 = 56\theta^3(1 - \theta)^5$$

b. Let's plot it!

```
n <- 8
y <- 3
theta <- seq(0, 1, length=200)
likelihood <- choose(n,y) * theta ^ y * (1 - theta) ^ (n - y)
plot(theta, likelihood, type='l')
```