HW1

STAT 651

RJ Cass

2026-01-16

1

Setup:

$P(p), P(n)$: Probability of positive/negative result

$P(C) = .001, P(nC) = .999$ Probability of cancer/no cancer

$P(p|C) = .98, P(n|nC) = .96 \implies P(n|C) = .02, P(p|nC) = .04$

$P(p) = P(p|C) * P(C) + P(p|nC) * P(nC) = .98(.001) + .04(.999) = 0.04094$

a)

$P(C|p) = \frac{P(p|C)P(C)}{P(p)} = \frac{.98(.001)}{0.04094} = .024$

b)

Given the likelihood of having cancer given a positive result is so low, I would suggest to the researchers that they either repeat the test multiple times on a subject (useful if test is cheap andfast) or find a way to reduce the chance of a positive result when the patient doesn't have cancer (given the current values, that will have a larger impact than increasing positive results if they do have cancer). Either way, I would advise them not to utilize the test as it currently is, as it is mostly likely going to cause panic in patients who actually have only a very small chance of actually having cancer.

2

a)

We will assume that we pick door 1, and that Monty will open door 2. The probability that the car is behind door 1 is 1/3. There are 3 possibilities for Monty opening door 2:

P1: P(Monty opens door 2 | Car behind door 1) = 1/2

P2: P(Monty opens door 2 | Car behind door 2) = 0

P3: P(Monty opens door 2 | Car behind door 3) = 1

Thus, P(Monty opens door 2) = P(B) = $P1 * 1/3 + P2 * 1/3 + P3 * 1/3 = 1/6 + 1/3 = 1/2$

Thus, P(Car behind door 1 | Monty opens door 2) = P(A|B) = $\frac{P(B|A)P(A)}{P(B)} = \frac{(1/2)(1/3)}{1/2} = 1/3$

This means the probability of the car being behind door 1 (the original choice) is only 1/3, where as if we switch the probability is 2/3.

b)

```r
set.seed(1337)
n <- 10000
no_switch <- 0
switch <- 0
for (i in 1:n) {
  car <- sample(c(1, 2, 3), 1)
  initial_guess <- 1
  if (car == initial_guess) {
    # If the car is the initial guess, just count that
    no_switch <- no_switch + 1
  } else {
    switch <- switch + 1
  }
}
```

I ran a simulation with 10^4 trials, which resulted in 3351 success when NOT switching, and 6649 successes when switching, roughly the same 1/3 vs. 2/3 ratio.

c)

For this, assume we will pick door 1. The below table indicates the result of switching or not in each situation (utility). The columns indicate which door the car is actually behind.

Switch	1	2	3
Yes	-$10	$200	$200
No	$200	-$10	-$10

Expected Utility Switching: -10(1/3) + 200(2/3) = 130

Expected Utility Not Switching: -10(2/3) + 200(1/3) = 60

3

a)

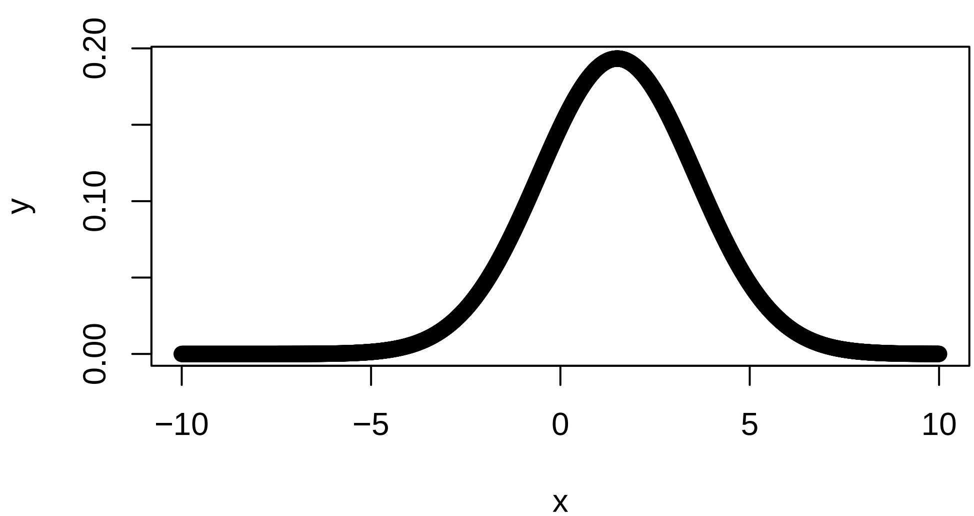
For $\sigma = 2$, what is marginal of y ($f(y)$)?

Multiply prior by each case: $f(y) = P(\theta = 1)f(y|\theta = 1) + P(\theta = 2)f(y|\theta = 2)$

$\sigma = 2 \implies f(y) = .5[N(y|1, 2^2)] + .5[N(y|2, 2^2)]$

A sketch (generated graph) of this is provided below.

```
x <- seq(-10, 10, .01)
y <- .5*dnorm(x, 1, 2) + .5*dnorm(x, 2, 2)
plot(x, y)
```



b)

Find $P(\theta = 1|y = 1)$, assuming $\sigma = 2$

$P(\theta = 1|y = 1) = \frac{f(y=1|\theta=1)P(\theta=1)}{P(y=1)} = \frac{.5f(y=1|\theta=1)}{P(\theta=1)f(y=1|\theta=1)+P(\theta=2)f(y=1|\theta=2)} = \frac{.5f(y=1|\theta=1)}{.5f(y=1|\theta=1)+.5f(y=1|\theta=2)}$

```r
f1 <- dnorm(1, 1, 2)
f2 <- dnorm(1, 2, 2)
p1 <- round(.5*f1 / (.5 * (f1 + f2)), 4)
```

Using R to do the calculation (dnorm(1, 1, 2), dnorm(1, 2, 2)): 0.5312

c)

As σ increases, the 2 potential distributions become less and less 'separable', meaning there's no real distinction between the two and the distribution for θ just looks like that of the prior (.5 for either case). As σ decreases, the possible distirbutions become more distinct meaning that given a value of y, it will be much more obvious which distirbution it came from (likliehood of it coming from a certain distribution approaches 1).

4

a)

Before the die is rolled, both people will assign the same probability (1/6) to a 6 being rolled. After the roll however, Person A will assign it either a 1 or 0 (they know the outcome), while person B still does not know and will continue assigning 1/6.

b)

Person A wouldn't really know much and would just probably just assign an equal chance to every country, or perhaps they would heavily weight towards their favorite country. Person B would base their probabilities off of the knowledge they have, and assign more probability to teams that have been performing well recently, etc.

5

a)

Impact on airlines: +$500 for each person who buys a ticket. -$3000 for each person over 100 who comes In all cases, Airline makes N * $500 per flight. For N <= 100, we don't care if anybody doesn't show: we won't have to return anything. For N > 100, the cost to the airline will be $3000 for each person who doesn't fit. We care when more than 100 people show up. Y $\sim$ Binom(N, θ). Number of people we need to pay fines for is outlined by: N - Y.

Utility function: $500 * N - 3000 * (N - Y)$

b)

```
N_list <- c(99:107)

calc_utility <- function(N, theta) {
  if (N <= 100) {
    N*500
  } else {
    bad_N <- 101:N
    prob_bad_N <- mapply(dbinom, x = bad_N, size = N, prob = theta)
    loss <- sum((bad_N - 100) * 3000 * prob_bad_N)
    500 * N - loss
  }
}

utility <- lapply(N_list, calc_utility, theta = .95)
cbind(N_list, utility)
```

```
      N_list utility
 [1,] 99     49500
 [2,] 100    50000
 [3,] 101    50483.13
 [4,] 102    50881.89
 [5,] 103    51067.62
 [6,] 104    50891.2
 [7,] 105    50248.52
 [8,] 106    49120.92
 [9,] 107    47569.02
```

c)

Given a θ of .95, the optimal value of N to maximize money made is 103 tickets.

d)

$p(y) = \int p(y|\theta)\pi(\theta)d\theta = \int Binom(N, \theta) * Beta(45, 2)d\theta = \int \binom{N}{y}\theta^y(1-\theta)^{N-y}\frac{\gamma(47)}{\gamma(45)\gamma(2)}\theta^{44}(1-\theta)^1$

$p(y) = \frac{\gamma(47)}{\gamma(45)\gamma(2)}\binom{N}{y}\int \theta^{y+44}(1-\theta)^{N-y+1}$

$\int \theta^{y+44}(1-\theta)^{N-y+1} = \int \theta^{(y+45)-1}(1-\theta)^{(N-y+2)-1} = \frac{\gamma(y+45)\gamma(N-y+2)}{\gamma(N-y+2+y+45)} = \frac{\gamma(y+45)\gamma(N-y+2)}{\gamma(N+47)}$

$p(y) = \frac{\gamma(47)\gamma(y+45)\gamma(N-y+2)}{\gamma(45)\gamma(2)\gamma(N+47)}\binom{N}{y} = \frac{46*45}{\gamma(2)\gamma(N+47)}\binom{N}{y}\gamma(y+45)\gamma(N-y+2)$

e)

```r
N_list <- c(99:107)

calc_y_prob = function(y, N) {
  choose(N, y) * gamma(y + 45) * gamma(N-y+2)
}

calc_utility_e <- function(N, theta) {
  if (N <= 100) {
    N*500
  } else {
    coef <- 46*45 / (gamma(2) * gamma(N + 47))
    bad_y <- 101:N
    prob_bad_y <- coef * mapply(calc_y_prob, y = bad_y, N = N)
    loss <- sum((bad_y - 100) * 3000 * prob_bad_y)
    500 * N - loss
  }
}

utility <- lapply(N_list, calc_utility_e, theta = .95)
cbind(N_list, utility)
```

```
     N_list utility
[1,] 99     49500
[2,] 100    50000
[3,] 101    50210.65
[4,] 102    50030.29
[5,] 103    49452.36
[6,] 104    48513.96
[7,] 105    47268.21
[8,] 106    45770.11
[9,] 107    44069.97
```

Based on this data, the airline should only sell 100 tickets.

f)

```r
N_list <- c(99:107)
n <- 100000

test_N <- function(N) {
  thetas <- rbeta(n, 45, 2)
  ys <- mapply(rbinom, n = 1, size = N, prob = thetas)
```

```r
  income <- N*500 - ifelse(ys > 100, (ys-100)*3000, 0)
  mean(income)
}

mean_income <- lapply(N_list, test_N)
cbind(N_list, mean_income)
```

```
      N_list mean_income
 [1,] 99     49500
 [2,] 100    50000
 [3,] 101    50209.96
 [4,] 102    50026.08
 [5,] 103    49456.7
 [6,] 104    48500.56
 [7,] 105    47262.51
 [8,] 106    45783.59
 [9,] 107    44052.61
```

This matches the calculated data above.

g)

```r
N_list <- c(99:107)
n <- 100000

test_N <- function(N) {
  thetas <- rbeta(n, 45, 2)
  ys <- mapply(rbinom, n = 1, size = N, prob = thetas)
  income <- N*500 - ifelse(ys > 100, (ys-100)*3000, 0)
  above_target <- income > 40000
  mean(above_target)
}

above_target_prop <- lapply(N_list, test_N)
cbind(N_list, above_target_prop)
```

```
      N_list above_target_prop
 [1,] 99     1
 [2,] 100    1
 [3,] 101    1
 [4,] 102    1
 [5,] 103    1
 [6,] 104    0.9073
 [7,] 105    0.90732
```

```
 [8,] 106    0.7826
 [9,] 107    0.64765
```

Based on this simulation, the airline can sell up to 105 tickets.

Homework Statements

Estimate of Time Taken

I estimate this homework took about 4 hours in total.

Disclosure of Resources Used

I used some online helps (Google, StackOverflow) for some help with debugging. I also used Wikipedia to verify my answer for the Beta-Binomial Distribution. Further collabed with a few classmates to verify answers for number 5.