Homework # 2

Elliott Pryor

9/1/2020

Problem 1

Statement

There are 20 trees in a plot and you must randomly choose 4 of the trees to take measurements on for a study.

1. In how many ways could you select the 4 trees?
2. Four of the trees are aspens and the rest are pines. In how many ways can you select 4 trees such that exactly two are aspen?
3. Now, assume you have to measure diameter at breast height for every tree in the plot. You decide it would be a good idea to do them in random order in case your measurements change as you become tired near the end of the day.
   a. In how many possible ways can you visit the 20 trees in the plot?
   b. If you randomly select the order in which you visit the 20 trees, what is the probability that you visit all four aspen trees back to back (with no pine trees in between)?

Solution

1. $\binom{20}{4} = \frac{20!}{16!4!} = \frac{20*19*18*17}{24} = 4845$
2. $\binom{16}{2} = \frac{16!}{14!2!} = \frac{16*15}{2} = 120$
3. a. $20! = 2.4329 * 10^{18}$
   b. (# Num of adjacent subsets of size 4) * (probability that a set of 4 trees is all aspen) = (Probability we visit all aspens in a row)

      $(16) * (\frac{4}{20} * \frac{3}{20} * \frac{2}{20} * \frac{1}{20}) = 0.024$

Problem 2

Statement

The eight-member Human Relations Advisory Board of Gainesville, Florida, considered the complaint of a woman who claimed discrimination based on gender, on the part of a local company. The board, composed of five women and three men, voted 5-3 in favor of the plaintiff, the five women voting in favor of the plaintiff, the three men against. The attorney representing the company appealed the board's decision by claiming gender bias on the part of the board members. Is gender bias a legitimate argument? Explain why or why not, and provide statistical evidence to support your reasoning

---

Solution

We consider the probability that out of 8 people, 5 female, 3 male, we end up with a 5:3 split exactly along the gender lines. We assume that there is no gender bias so each individual is equally likely to vote yes and no. We order the individuals: $f_1, f_2, f_3, f_4, f_5, m_1, m_2, m_3$. The probability that each individual votes yes or no $P(Y_j) = P(N_j) = 0.5$. In this scenario we have: $Y_{f_1}, Y_{f_2}, Y_{f_3}, Y_{f_4}, Y_{f_5}, N_{m_1}, N_{m_2}, N_{m_3}$. This is the only possible configuration of votes. The probability this occurs is $1/2 * 1/2 * 1/2 * 1/2 * 1/2 * 1/2 * 1/2 * 1/2 = (1/2)^8 = 1/256 = 0.0039$. This is exceedingly unlikely to occur randomly. So yes gender bias is a legitimate argument.

Problem 3

Statement

If m students show up at random in a classroom, what is the probability that at least two will have the same birthday? You may assume, there are 365 days in every year (ignore leap years) and everyday of the year is equally likely to be a birthday.

1. A function to calculate the probability for a specific value of m, is provided in birthday.R under content. Use the function to compute the probability for different values of m between 2 (you'd need at least two people in a room to have a shared birthday) and 100 and plot the results. Discuss what your plot shows.

2. Use your results to answer the following question: How many people need to be in the room to make it more likely than not that at least two people share the same birthday? That is, what is m such that $P(A = \text{at least two people share the same birthday}) > 0.5$

Solution

1. The provided function takes a value of m and computes the probability of at least 2 of the m students sharing a birthday.

```
bday_fun <- function(m){
  # compute the probability of complement of A on log scale
  l_prob_a_c <- lfactorial(365) - lfactorial(365-m) - m*log(365)
  # find probability of A
  prob_a <- 1- exp(l_prob_a_c)
  # return probability of A
  return(prob_a)
}
```

We can then apply the function to plot the results

```
mVals <- 2:100
probs <- bday_fun(mVals)
plot(x = mVals, y = probs, xlab = "# of people in the room",
     ylab = "Probability of at least one shared bday (P(A))")
```

Figure 1 shows the probability that two people will share a birthday. This results in a roughly logisistic pattern. With few people the probability is very low that two people will share a birthday. The probability that two people will share a birthday increases very rapidly between [10, 40] after which it remains close to 1.

2. We can compute how many people it takes until we expect at least one shared birthday.

```
greatProb <- probs > 0.5
indx <- min(which(greatProb, arr.ind = TRUE))
# by our construction of probs, index 1 --> 2 people, index 2 -->3 people, index i --> i+1 people
people <- indx + 1
people
```

```
## [1] 23
```

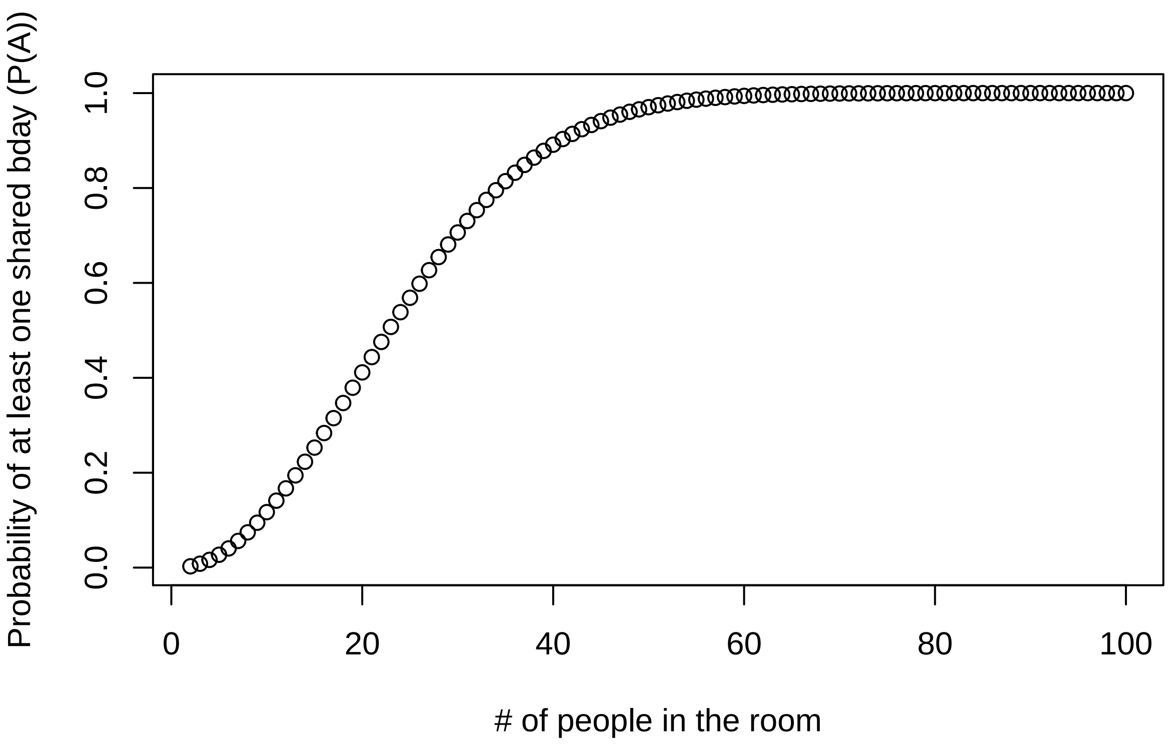

Figure 1: Probability that two people will share a birthday with varying numbers of people

So the minimum number of people required, such that the probability of at least two people sharing a birthday is greater than 50%, is 23 people.