

S&DS 238 Problem Set #1

Shaun Radgowski

September 11, 2020

Problem 1

(1a)

The probability that any given person has a different birthday than you is the number of days that are not your birthday divided by the total number of days, which is $364/365$. The probability that k people do *not* share your birthday is thus:

$$\left(\frac{364}{365}\right)^k$$

Using the complement rule, the probability that *at least one* person of the k people shares your birthday is then:

$$p(k) = 1 - \left(\frac{364}{365}\right)^k$$

(1b)

To find the smallest number of people k^* such that $p(k^*) \geq 0.9$, let's first calculate how many people are required for the probability to equal 0.9. That would be:

$$0.9 = 1 - \left(\frac{364}{365}\right)^k \implies \left(\frac{364}{365}\right)^k = 0.1 \implies k = \log_{\frac{364}{365}} 0.1 = 839.2917$$

From there, k^* is simply the next-largest integer, or **840**.

(1c)

To run this simulation and print out the fraction of repetitions with at least one match of my birthday, we can use the following code:

```
k <- 840 # Number of other people
n <- 10000 # Number of simulations to be run
match_counts <- rep(NA, n)
total_success <- 0 # Number of simulations with at least 1 other appearance of my birthday
for (i in 1:n) {
  me <- sample(1:365, size = 1) # My own random birthday
  others <- sample(1:365, size = k, replace = T) # Their random birthdays
  no_matches <- sum(others == me)
  match_counts[i] <- no_matches
}
```

```

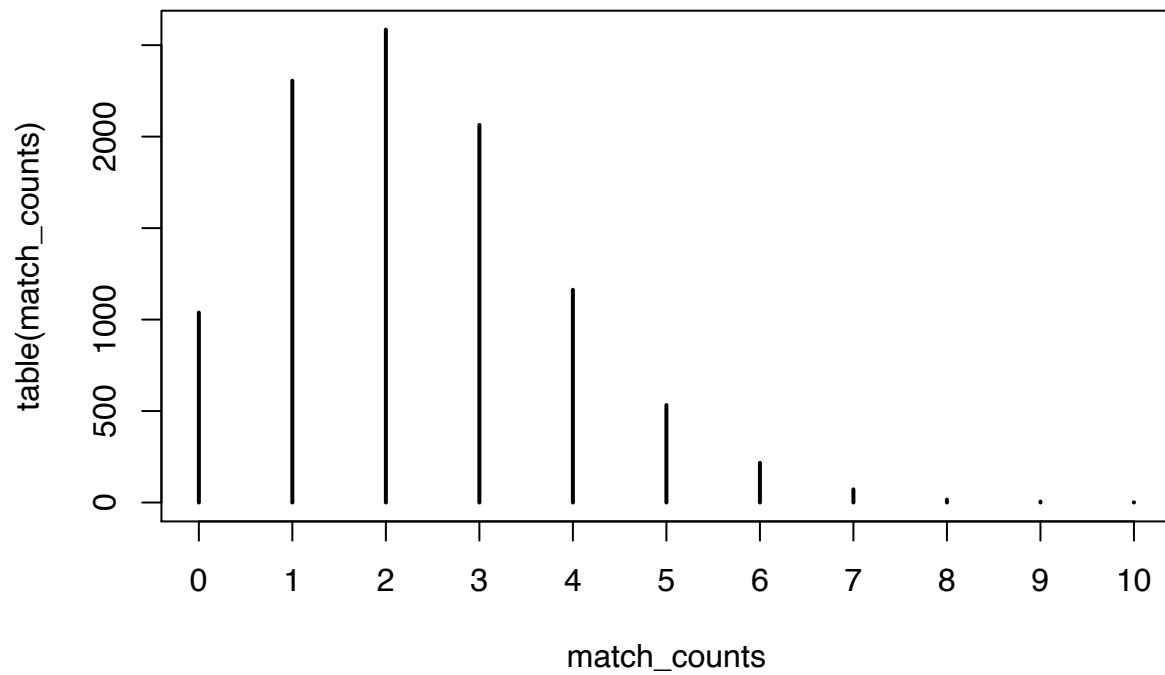
if (no_matches > 0) {
  total_success <- total_success + 1
}
}

total_success/n

```

```
## [1] 0.8962
```

```
plot(table(match_counts))
```



Problem 2

(2a)

To create the first vector (20 through 1, decreasing by 1), we could use the code:

```
20:1
```

```
## [1] 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
```

To create the second vector (strings “x1” through “x20”), we could use the code:

```
paste("x", 1:20, sep = "")
```

```
## [1] "x1" "x2" "x3" "x4" "x5" "x6" "x7" "x8" "x9" "x10" "x11" "x12"
## [13] "x13" "x14" "x15" "x16" "x17" "x18" "x19" "x20"
```

To create the third vector (20 3s, 10 5s, and 30 2s), we could use the code:

```
c(rep(3, 20), rep(5, 10), rep(2, 30))
```

```
## [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 5 5 5 5 5 5 5 5 5 5 2 2 2 2 2 2 2 2
## [39] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```

(2b)

To calculate the sum

$$\sum_{i=21}^{51} 2i^3 + 3i^2$$

without a loop, we could use the code:

```
index <- seq(21, 50)
sum(2 * (index^3) + 3 * (index^2))
```

```
## [1] 3283215
```

To calculate the same sum *with* a loop, we could then use the code:

```
total <- 0
for (i in 21:50) {
  total <- total + 2 * (i^3) + 3 * (i^2)
}
total
```

```
## [1] 3283215
```

Problem 3

(3a)

There are 6 varieties for the type of pair (one for each possible die roll, 1-6), which can be multiplied by the number of ways to arrange that pair among the 5 dice. The remaining three dice must all be different values, thus we are picking 3 distinct values from the 5 remaining choices. Once those values are picked, we must multiply by the total number of ways to arrange those three distinct values among the three other dice, which is 3!. The probability is thus the total number of one-pair hands divided by the total number of possible hands.

$$6 * \frac{\binom{5}{2} * \binom{5}{3} * 3!}{6^5} = \frac{\binom{5}{2} * \binom{5}{3}}{6^3} = \frac{100}{216} = 0.4630$$

(3b)

There are 6 varieties of the type of 3-of-a-kind (one for each possible die roll, 1-6), leaving 5 varieties for the pair, for $6 \cdot 5 = 30$ total options for full houses. The total number of full houses is 30 times the number of unique ways to arrange three of the one number times the number of unique ways to arrange two of another number. The probability is thus the total number of full houses divided by the total number of possible hands.

$$30 * \frac{\binom{5}{3} * 1}{6^5} = \frac{300}{7776} = 0.0386$$

(3c)

Using the same logic as before, there are 6 varieties for the first pair and 5 varieties for the second pair. But, these two pairs are interchangeable, so that means there are *6 Choose 2* total options for two-pair hands. Each of these varieties has the same probability due to symmetry, so the total number of two-pair hands is 30 times the number of unique ways to arrange the first pair among the five dice, times the number of unique ways to arrange the second pair among the remaining three dice, times the remaining 4 possibilities for the final die (it can be any number *except* the values from the two pairs). The probability is thus the total number of two-pair hands divided by the total number of possible hands.

$$\binom{6}{2} * \frac{\binom{5}{2} * \binom{3}{2} * 4}{6^5} = \frac{1800}{7776} = 0.2315$$

Problem 4

To design a simulation that rolls 5 dice and keeps track of the proportion of two-pair hands that come up, we can use the following code:

```
n = 10000 # Number of simulations
total = 0 # Number of two-pair hands
for (i in 1:n) {
  dice <- sample(1:6, size = 5, replace = T)
  # Count how many of each value appear; if there are 2 2s, it's a success
  df <- as.vector(table(dice))
  if (length(which(df == 2)) == 2) {
    total <- total + 1
  }
}

total/n
```

```
## [1] 0.2286
```

It seems the average proportion is around the predicted 0.2315.

Problem 5

(5a)

By the Third Axiom of Probability, the three scenarios in the first line are all disjoint, so the probability of a union of these events is the sum of their probabilities.

$$A \cup B = (AB^C) \cup (A^C B) \cup (A \cap B) \quad P(A \cup B) = P(AB^C) + P(A^C B) + P(A \cap B)$$

Now, by the Second Axiom of Probability, the total probability across all events in a sample space must equal 1. This means that for events A and B, the following must be true:

$$A = (AB) \cup (AB^C), \quad P(A) = P(A \cap B) + P(A \cap B^C)$$

$$P(A \cap B^C) = P(A) - P(A \cap B), \quad P(A^C \cap B) = P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \Delta B) = P(A \cup B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$$

(5b)

According to Bayes' Theorem, for any events A and B:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)}{P(B)}P(A)$$

The fractional term of $\frac{P(B|A)}{P(B)}$ we know to be greater than 1, because of what we were told in the problem statement:

$$P(B|A) > P(B)$$

Thus, it is clear that $P(A|B) = kP(A)$, where k is a positive value greater than 1. This proves that $P(A|B) > P(A)$, and that the occurrence of B increases the occurrence of A more likely.

(5c)

i.

If A and B are disjoint, then:

$$P(B) = P(A \cup B) - P(A) = 0.9 - 0.6 = 0.3$$

ii.

If A and B are independent, then:

$$P(AB) = P(A)P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A)P(B)$$

$$P(B)(1 - P(A)) = P(A \cup B) - P(A)$$

$$P(B) = \frac{P(A \cup B) - P(A)}{1 - P(A)}$$

$$= \frac{0.9 - 0.6}{0.4} = 0.75$$