

## Problem Set 1

Submitted by: Giulia Petrilli

Group: Henry Baker, Jennifer Estigene

### Question 1

- (a) This is the probability that we sample 60 *unique* units from a population of 600, while sampling with replacement, where order does not matter.  
This is a version of the birthday problem.

$$P(\text{No Match}) = \frac{\text{Number of ways to not repeat}}{\text{Number of total possibilities}} \quad (1)$$

Where *Population*,  $n = 600$ , *choices*,  $k = 60$

*Numerator*: this is the number of successful possibilities. To find 60 unique respondents here we sample with replacement, unordered:  $n(n-1) \dots (n-k+1)$ .

*Denominator*: this is the sample space. We sample with replacement, ordered:  $n^k$

The probability function is:

$$P(60 \text{ unique respondents}) = \frac{600}{600} \cdot \frac{599}{600} \cdots \frac{541}{600} \quad (2)$$

Using R, this calculates to a probability of  $\approx 0.0472$ .

R Code:

```
p_unique <- 1
for(i in 0:59)
  p_unique <- p_unique * (600 - i) / 600
round(p_unique, 4)
```

- (b) This set up is the same as (a), but this time without replacement (and again, order does not matter). To count for no replacement, unordered sampling we use the *Binomial Coefficient*  $\binom{n}{k}$ . The binomial coefficient counts the number of possible (unordered) subsets of size  $k$ , for a set sized  $n$ .

Since the binomial coefficient is calculated as follows:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

We plug in the numbers:

$$\binom{600}{60} = \frac{600!}{60!(540)} \quad (3)$$

Using R we calculate this to construct a sample of 60 from a population of 600, when order does not matter as  $\approx 2.774267 \times 10^{83}$ .

```
R Code:
result_b <- choose(600, 60)
result_b
```

- (c) This is the double tagged 'story' for which we use *Hypergeometric distribution*, where we set being a student to a 'success'.

Setting the r.v to  $X = k$ , the Hypergeometric distributions answers the question: what is the likelihood of finding  $k$  'successes' (students) in a sample size of  $n$ , given that in the actual population of size  $N$  there are  $K$  number of 'successes' (students). This gives us:

- Population Size,  $N = 600$
- Number of 'successes' in the population,  $K = 400$
- Sample size,  $n = 60$
- Then finally, we set the number of successes in the sample for which we want to find the associated probability for,  $k = 60$

The Hypergeometric Distribution PMF:

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}} \quad (4)$$

$$P(X = 60) = \frac{\binom{400}{60} \cdot \binom{200}{0}}{\binom{600}{60}} \quad (5)$$

$\binom{200}{0} = 1$  (only one way to choose zero units). So, this simplifies to:

$$P(X = 60) = \frac{\binom{400}{60}}{\binom{600}{60}} \quad (6)$$

Now we have to calculate a series of binomial coefficients. Substituting these binomial coefficients back into the equation, we have:

$$P(X = 60) = \frac{\frac{400!}{60! \cdot 340!}}{\frac{600!}{60! \cdot 540!}} \quad (7)$$

Using R to calculate, the probability of the survey sample containing only students  $\approx 5.44 \times 10^{-12}$ .

```
R Code:
result_c <- choose(400, 60) / choose(600, 60)
result_c
```

- (d) To calculate the total number of ways of constructing this sample, first we calculate the ways to construct the samples Within each blocks (without replacement, unordered - the *Binomial Coefficient*). Then we use the multiplication rule between blocks to get

to the total number of possibilities.

$$\text{Block A: } n = 400, k = 40 \quad (8)$$

$$\text{Block B: } n = 50, k = 5 \quad (9)$$

$$\text{Block C: } n = 150, k = 15 \quad (10)$$

$$(11)$$

$$\binom{400}{40} \cdot \binom{50}{5} \cdot \binom{150}{15} = \frac{400!}{40!(360)!} \cdot \frac{50!}{5!(45)!} \cdot \frac{150!}{15!(135)!} = 6.78 \times 10^{81} \quad (12)$$

R Code:

```
result_d <- choose(400, 40) * choose(50, 5) * choose(150, 15)
result_d
```

- (e) We (i) first choose 10 people from 600 and then (ii) arrange them in a circular manner. Choosing 10 people from 600 we use the binomial coefficient Where  $n = 600, k = 10$  gives us  $\binom{600}{10}$ .

Next, we use factorial permutation to arrange them in a circular manner we calculate by  $(10 - 1)!$ . (NB circular arrangements are counted as  $\frac{n!}{n} = (n - 1)!$ . i.e distinct from arranging units along a straight line.)

Thus, the total number of ways to construct the focus group, considering different seating arrangements as distinct, is:

$$\binom{600}{10} \cdot 9! = 5.61 \times 10^{26}$$

As we will get into in more depth in 1(f), our answer above rests on an assumption that we are considering the fact the table is circular. In the below R code we also calculate for if we were to relax that assumption (and thus to calculate the number of linear permutations as  $10!$ , giving us  $5.61 \times 10^{27}$ ).

R Code:

```
result_e_1 <- factorial(9) * choose(600, 10)
result_e_1
```

```
result_e_2 <- factorial(10) * choose(600, 10)
result_e_2
```

- (f) Here we are interested in counting the number of ways to obtain a so-called successful arrangement, over the count of the number of total arrangements.

$$P = \frac{\text{Successful Arrangements}}{\text{Total Arrangements}}$$

We provide the probability of a successful arrangement both in a linear and in a circular arrangement. The solution will exhaustively express the probability that all five students are seated next to each other in both scenarios.

## 1 Linear permutation

*Numerator:*

We seek to find out what is the probability that five students in our group of 10 are seated next to each other

Since we are interested in the 5 students together and not singularly, we consider them as a group of 1.

Number of other individual group members (5) + the students' group (1) = ((6!)

For each student within that block of 5 students, there are ((5)! possible ways they could arrange their seating

Using the multiplication rule, the number of favorable arrangements is:  $6! \times 5!$

*Denominator*

We are interested in the total number of arrangements. Since it's a linear arrangement of 10 people, the denominator will imply be  $10!$ . Thus, the probability that all 5 students are seated next to each other is:

$$\frac{6! \times 5!}{10!}$$

The probability that all five students are seated next to each other in a linear fashion is  $\approx 0.02381$ .

*R Code:*

```
result_f_1 <- (factorial(6) * factorial(5)) / (factorial(10))
print(round(result_f_1, 5))
```

## 2 Circular permutation

Here, we are looking at the probability that five students in our group of 10 are seated next to each other in a circular arrangement

*Numerator:*

The permutations for the seating order in a circular fashion are  $((6!-1!))$

The permutations for each student to sit within that block is the same as in the linear arrangement,  $((5!))$

Using the multiplication rule, the number of favorable arrangements is:  $5! \times 5!$

*Denominator*

Since it's a circular arrangement of 10 people, we have  $(10 - 1)! = 9!$ . Thus, the probability that all 5 students are seated next to each other is:

$$\frac{5! \times 5!}{9!}$$

The probability that all five students are seated next to each other in a linear fashion is  $\approx 0.03968$ .

*R Code:*

```
result_f_2<- (factorial(5) * factorial(5)) / (factorial(9))
print(round(result_f_2, 5))
```

*There is an extended version of this where we also explicitly calculate the probability of 5 students sitting next to each other in the subset, considering sampling from the original population ratios.*

Linear arrangement

$$P = \frac{\binom{200}{5} \binom{400}{5} \cdot 6! \cdot 5!}{\binom{200}{5} \binom{400}{5} 10!}$$

Circular Arrangement

$$P = \frac{\binom{200}{5} \binom{400}{5} \cdot 5! \cdot 5!}{\binom{200}{5} \binom{400}{5} 9!}$$

*As expected, the probabilities are the same as if we did not insert  $\binom{400}{5}$  and  $\binom{200}{5}$  in both the Numerator and the Denominator. To check, we used the R codes `result_f_3` and `result_f_4`.*

R Code:

```
result_f_1 <- (factorial(6) * factorial(5)) / (factorial(10))
print(round(result_f_1, 5))
```

```
result_f_2<- (factorial(5) * factorial(5)) / (factorial(9))
print(round(result_f_2, 5))
```

```
result_f_3 <- (factorial(5) * factorial(6)*
choose(200, 5)* choose(400, 5)) /
(factorial(10) *choose(200, 5)* choose(400, 5))
round(result_f_3, 6)
```

```
result_f_4 <- (factorial(5) * factorial(5)*
choose(200, 5)* choose(400, 5)) / (factorial(9)
*choose(200, 5)* choose(400, 5))
round(result_f_4, 6)
```

## Question 2

- (a) Let prevalence of disease throughout the population be 0.01 (expressed in probability terms:  $P(D) = 0.01$ )  
 Let the test's sensitivity (*true positive*) be  $P(T|D) = 0.95$ .  
 Let the test's specificity (*true negative*) be  $P(T^c|D^c) = 0.95$ .

First, the accuracy of a test expressed as a function of sensitivity and specificity:

$$Accuracy = (sensitivity * prevalence) + (specificity * (1 - prevalence)) \quad (13)$$

Where *prevalence* represents the proportion of the population that actually has the condition being tested for. If we plug the numbers in:

$$Accuracy_{Test} = (0.95 * 0.01) + (0.95 * (1 - 0.01)) \quad (14)$$

Accuracy of 0.95 intuitively makes sense given that sensitivity and specificity are both the 0.95. As accuracy is just a weighted sum function of these two parameters, if they are the same then the final measure of accuracy will necessarily reflect that.

- (b) Company B's test is indeed more accurate. This is intuitively the case as it is completely *specific* (ie it *always* identifies true negatives), and in this instance, given the low prevalence of the disease in the population), the function for accuracy weights specificity a good deal more than sensitivity (with a weight of 0.99 to 0.01). Thus even though B's test is completely *insensitive* (indeed it's sensitivity is 0), its accuracy is greater. Plugging the numbers in:

$$Accuracy_B = (0 * 0.01) + (1 * (1 - 0.01)) \quad (15)$$

$$= 0.99 \quad (16)$$

- (c) Because company B's test is completely insensitive it is unable to detect the presence of the disease even when it is there. While company A's test produces some false positives (at a rate of 0.05), it is nevertheless sensitive to the disease's presence. It is thus a valid measurement device (unlike B's test). This is crucial in the context of public health responses as A's test allows for detection and treatment, whereas B's test will only provide false negatives when it is used to test a patient who does in fact carry the disease. This is a dangerous situation, leading to worse net outcomes as a population level, as well as an individual level.
- (d) Given the accuracy of company B's test at 0.99, company A must have an accuracy greater than 0.99.

- if sensitivity = specificity - setting both values to  $x$ :

$$0.99 < (x * 0.01) + (x * 0.99) \quad (17)$$

$$0.99 < 0.01x + 0.99x \quad (18)$$

$$0.99 < x \quad (19)$$

$$x > 0.99 \quad (20)$$

Both would have to be above 0.99. This intuitively makes sense for the same reason mentioned above; if accuracy is the sum of the weighted values of specificity and sensitivity, and if they are both necessarily equal, then they would *both* need to be above the 0.99 threshold.

- if sensitivity = 1 - setting specificity value to  $x$

$$0.99 < (1 * 0.01) + (x * 0.99) \quad (21)$$

$$0.99 < 0.01 + 0.99x \quad (22)$$

$$0.99x > 0.98 \quad (23)$$

$$x > 0.98/0.99 \quad (24)$$

$$x > 0.989898989899 \quad (25)$$

Specificity must be greater than 0.99 (rounded to 2 d.p.), otherwise above 0.98 recurring.

- if specificity = 1 - setting sensitivity value to  $x$ :

$$0.99 < (x * 0.01) + (1 * 0.99) \quad (26)$$

$$0.99 < 0.01x + 0.99 \quad (27)$$

$$0.99 - 0.99 < 0.01x \quad (28)$$

$$0 < 0.01x \quad (29)$$

$$x > 0 \quad (30)$$

Sensitivity must be greater than 0; there must just be some sensitivity (the test must be not completely insensitive). This also makes intuitive sense; Company B's test is completely specific (ie = 1), but completely insensitive (ie = 0), then if Company A also has a completely specific test, all Company A needs to do to better Company B's test in terms of accuracy is to have a sensitivity level greater than 0.

### Question 3

- (a) If  $P = \text{Heads}$ , we calculate the probability of success when switching under the normal Monty hall. If  $(1 - P) = \text{Tails}$ , on the other hand, we calculate the probability of success when switching after Monty simply opened the door chosen, regardless it concealing the Car or not. To solve this exercise, we employ Bayes rule and LOTP.

$Pr(C3|M2)$  for  $H, T$

$$\Pr(C_3 | M_2) = P \cdot \Pr(C_3 | M_2, H) + (1 - P) \cdot \Pr(C_3 | M_2, T) \quad (31)$$

$$\Pr(C_3 | M_2, H) = \frac{P(M_2 | C_3) P(C_3)}{P(M_2)} \quad (32)$$

$$\frac{P(M_2 | C_3) P(C_3)}{P(M_2 | C_1) P(C_1) + P(M_2 | C_2) P(C_2) + P(M_2 | C_3) P(C_3)} \quad (33)$$

$$\frac{(1) \left(\frac{1}{3}\right)}{\left(\frac{1}{2}\right) \left(\frac{1}{3}\right) + (0) \left(\frac{1}{3}\right) + (1) \left(\frac{1}{3}\right)} \quad (34)$$

$$\text{When Heads} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} \quad (35)$$

$$(1 - p) \Pr(C_3 | M_2, T) = \frac{P(M_2 | C_3) P(C_3)}{P(M_2)} \quad (36)$$

$$\frac{P(M_2 | C_3) P(C_3)}{P(M_2 | C_1) P(C_1) + P(M_2 | C_2) P(C_2) + P(M_2 | C_3) P(C_3)} \quad (37)$$

$$\frac{\frac{1}{6}}{\left(\frac{1}{2}\right) \left(\frac{1}{3}\right) + (0) \left(\frac{1}{3}\right) + \left(\frac{1}{2}\right) \left(\frac{1}{3}\right)} \quad (38)$$

$$\text{When Tails} = \frac{1}{2} \quad (39)$$

In H,  $P(M2|C3)$  is 1, hence probability of success  $P(C3|M3)$  is  $2/3$

In T,  $P(M2|C3)$  is  $1/2$ , hence probability of success  $P(C3|M3)$  is  $1/2$

(Explanation)  $P(M2|C3)$  is 1 in Heads because if the Car is behind Door 3 and the contestant chose Door 1, Monty has no choice but to open Door 2.  $P(M2|C3)$  is  $1/2$  in Tails because if the car is behind Door 1 and the contestant also chose Door 1, Monty has a choice to open either Door 2 or Door 3. So, the probability he chooses Door 2 is  $1/2$  in Tails.

We compute the probability that as Door 1 is chosen, Monty opens Door 2 and the contestant switches to Door 3 as follows:

$$\left(p \times \frac{2}{3}\right) + \left(\frac{1}{2} \times (1 - p)\right) \quad (40)$$

$$\left(\frac{2p}{3}\right) + \left(\frac{1-p}{2}\right) \frac{2p}{3} \times \frac{2}{2} = \frac{4p}{6} \quad (41)$$

$$\frac{1-p}{2} \times \frac{3}{3} = \frac{3-3p}{6} \quad (42)$$

$$\frac{4p + 3 - 3p}{6} \quad (43)$$



(b) R code

In Heads world the probability of winning the car is significantly higher than when they stick with their first choice. However, in Tails world, the probability of success when switching is 0.5. The result in this simulation,  $\approx 0.503$ , show that there is a slightly higher advantage in switching.

```
set.seed(123)

n <- 10000
p <- 0.5
success <- 0

for (i in 1:n) {
  coin <- ifelse(runif(1) < p, "H", "T")
  choice <- 1
  monty <- if (coin == "H") ifelse(choice == 1, sample(2:3, 1), 3)
  else sample(setdiff(1:3, choice), 1)
  remaining <- setdiff(1:3, c(choice, monty))
  if (remaining == 3) success <- success + 1
}

prob_success <- success / n
cat("Simulated Probability of Success\n",
    "(Switching to Door 3):", prob_success, "\n")
```

## Question 4

- (a)  $X$  is a discrete r.v. where its values  $x$  represent levels of the game.  $X$  is a variable representing the highest level a player reaches. So,  $X = 1$ , for example is the event that the player does not pass level 1, the probability of which is  $1 - p_1$ .

PMF of  $X$

$$P(X = x) = \begin{cases} 1 - p_1 & \text{if } x = 1 \\ p_1 \cdot (1 - p_2) & \text{if } x = 2 \\ p_1 \cdot p_2 \cdot (1 - p_3) & \text{if } x = 3 \\ p_1 \cdot p_2 \cdot p_3 \cdot (1 - p_4) & \text{if } x = 4 \\ p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot (1 - p_5) & \text{if } x = 5 \\ p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5 \cdot (1 - p_6) & \text{if } x = 6 \\ p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5 \cdot p_6 & \text{if } x = 7 \end{cases} \quad (44)$$

- (b) To meet the necessary conditions for a valid Probability Mass Function (PMF), two criteria must be satisfied:
- (1.) Non-negativity, where the probability  $P(X = x) > 0$  when  $x = x_j$  for some  $j$ , and  $p_x(x) = 0$  otherwise.
  - (2.) The sum of all probabilities must equal to 1.

(Requirement 1) Since the probability of passing through each level can be conceived of as the probability given by an independent Bernoulli trial, each  $p_j$  must necessarily fall within the range of 0 to 1 inclusive. (in fact just by their being probabilities their valued by definition lie between 0 and 1 inclusive). Given that each term in the PMF represents a product of probabilities or  $(1 - \text{probability})$ , it is necessarily the case that each term is also non-negative.

(Requirement 2) We can aggregate the individual probabilities for each  $x$  as follows:

$$\begin{aligned} & (1 - p_1) + (p_1 \dots p_6) + \sum_{x=2}^6 p_1 \dots p_{x-1} (1 - p_x) \\ &= 1 - p_1 + p_1 - p_1 p_2 + p_1 p_2 - p_1 p_2 p_3 + p_1 p_2 p_3 - p_1 p_2 p_3 p_4 + p_1 p_2 p_3 p_4 \\ & \quad - p_1 p_2 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5 - p_1 p_2 p_3 p_4 p_5 p_6 + p_1 p_2 p_3 p_4 p_5 p_6 \\ &= 1 \end{aligned}$$

All the terms from  $p_1$  to  $p_1 p_2 \dots p_6$  cancel out, leaving only the 1; the sum of the probabilities in the PMF equal 1.

R Code:

```
check_PMF_validity <- function(p1, p2, p3, p4, p5, p6)
{p_values <- c(1 - p1, p1 * (1 - p2), p1 * p2 * (1 - p3),
p1 * p2 * p3 * (1 - p4), p1 * p2 * p3 * p4 * (1 - p5),
p1 * p2 * p3 * p4 * p5 * (1 - p6), p1 * p2 * p3 * p4 * p5 * p6)
total <- sum(p_values)
is_valid <- abs(total - 1) < 1e-9
return(list(sum = total, is_valid = is_valid))}

# Test the function with random values
p1 <- runif(1)
```

```

p2 <- runif(1)
p3 <- runif(1)
p4 <- runif(1)
p5 <- runif(1)
p6 <- runif(1)

cat("p1:", p1, "\np2:", p2, "\np3:", p3, "\np4:", p4,
    "\np5:", p5, "\np6:", p6, "\n")

result <- check_PMF_validity(p1, p2, p3, p4, p5, p6)
cat("Sum:", result[["sum"]], "\nIs Valid:", result[["is_valid"]], "\n")

```

- (c) If the probability of getting to the next level is the same for all levels then the more concise PMF reads as follows:

$$P(X = x) = \begin{cases} 1 - p & \text{if } x = 1 \\ p(1 - p) & \text{if } x = 2 \\ p \cdot p(1 - p) & \text{if } x = 3 \\ p \cdot p \cdot p(1 - p) & \text{if } x = 4 \\ p \cdot p \cdot p \cdot p(1 - p) & \text{if } x = 5 \\ p \cdot p \cdot p \cdot p \cdot p(1 - p) & \text{if } x = 6 \\ p \cdot p \cdot p \cdot p \cdot p & \text{if } x = 7 \end{cases}$$

Compacted version:

$$P(X = x) = \begin{cases} p^{x-1}(1 - p) & \text{if } 1 \leq x \leq 6 \\ p^6 & \text{if } x = 7 \end{cases}$$

- (d) Set  $X$  as the number of levels failed before beating 3 levels, with the probability of success at any level as  $p$ . There are 3 possible scenarios (support of  $X$  is  $1 : 3$ ;  $x = 1 : 3$ ). You can fail 0 times, meaning you beat the first 3 levels successfully; you can fail 1 time before beating 3 levels; you can fail 2 times before beating 3 levels; you can fail 3 times before beating 3 levels.

Additionally there are multiple arrangements of failing-succeeding (for  $X = 1$ ,  $X = 2$ ).

$X = 1$  can occur in sequences: FSS, SFS, SSF = 3 possible permutations.

$X = 2$  can occur in sequences: FFS, FSF, SFF = 3 possible permutations.

*(Whereas for  $X = 0$ ,  $X = 3$ , there's only one possibility of arranging the sequences)*

The PMF of  $X$  is as follows:

$$P(X = x) = \begin{cases} 1 \cdot p^3 \cdot (1 - p)^0 & \text{if } x = 0 \\ 3 \cdot p^2 \cdot (1 - p)^1 & \text{if } x = 1 \\ 3 \cdot p^1 \cdot (1 - p)^2 & \text{if } x = 2 \\ 1 \cdot p^0 \cdot (1 - p)^3 & \text{if } x = 3 \end{cases}$$

I wrote out the above in full for full clarity as to the pattern of development. But this

PMF can be rewritten without the redundant terms as:

$$P(X = x) = \begin{cases} p^3 & \text{if } x = 0 \\ 3p^2 \cdot (1 - p)^1 & \text{if } x = 1 \\ 3p^1 \cdot (1 - p)^2 & \text{if } x = 2 \\ (1 - p)^3 & \text{if } x = 3 \end{cases}$$

If we return to the piece wise function PMF *with* the redundant terms, we can clearly see a pattern emerge, which allows us to write the PMF as a more concise function. Specifically, the number of ways we can arrange successful and failure Bernoulli trials is a permutation calculation. When all trials are either successes or failures ( $X = 0, X = 3$ ), then there is only one way to arrange them - as given by  $\binom{3}{3}$  and  $\binom{3}{0}$ . Whereas when one or two trials result in successes (or failures; they are complements of each other) both have 3 possible arrangement permutations - as is given by  $\binom{3}{1}$  and  $\binom{3}{2}$ . With this, we can now rewrite out the PMF of  $x$  using binomial coefficients.

$$P(X = x) = \binom{3}{x} \cdot p^{3-x} \cdot (1 - p)^x \quad (45)$$

- (e) All of this is just deriving the Negative Binomial distribution (which models the probability distribution of number of failures before achieving a fixed number of successes). Here we are interested in the number of failures ( $X$ ) before 3 successes, with the probability of success as  $p$ .

The generic PMF of a given Negative Binomial Distributions is:

$$P(X = k) = \binom{k + r - 1}{k} p^r (1 - p)^k$$

Where:

- $k$  is the number of failures before the  $r$ -th success.
- $p$  is the probability of success on any given trial.
- $r$  is the number of successes we are (usually) interested in.

The binomial distribution and the negative binomial distribution model sequences of Bernoulli trials, but they are used to answer different types of questions.

- The *Binomial Distribution* gives the probability of  $k$  successes in  $n$  trials,
- The *Negative Binomial Distribution* gives the probability of  $k$  failures before  $r$  given successes.

We use the negative binomial when we want to find the probability of a specified number of failures/successes occurring before a predetermined number of successes/failures. *+ the negative binomial distribution assumes that the sequence of trials stops after the  $r$ th success.*

## 1. ANNEX - R CODES

## (a) Question 1:

R Code for Question 1a:

```
p_unique <- 1
for(i in 0:59)
  p_unique <- p_unique * (600 - i) / 600
round(p_unique, 4)
```

R Code for Question 1b:

```
result_b <- choose(600, 60)
result_b
```

R Code for Question 1c:

```
result_c <- choose(400, 60) / choose(600, 60)
result_c
```

R Code for Question 1d:

```
result_d <- choose(400, 40) * choose(50, 5) * choose(150, 15)
result_d
```

R Code for Question 1e:

```
result_e_1 <- factorial(9) * choose(600, 10)
result_e_1
```

```
result_e_2 <- factorial(10) * choose(600, 10)
result_e_2
```

R Code for Question 1f:

```
result_f_1 <- choose(400, 5) * choose(200, 5)
result_f_1
```

```
result_f_2 <- (factorial(5) * choose(400, 5)
* factorial(5) * choose(200, 5)) / (factorial(9))
round(result_f_2, 5)
```

```
result_f_3 <- (factorial(5) * choose(400, 5)
* factorial(5) * choose(200, 5)) / (factorial(10))
round(result_f_3, 6)
```

## (b) Question 2:

R Code for Question 2a:

```
# Your R code for Question 2a here
```

(c) Question 3:

R Code for Question 3b:

# Your R code for Question 3b here

R Code:

```
result_f_1 <- (factorial(6) * factorial(5)) / (factorial(10))
print(round(result_f_1, 5))
```

```
result_f_2<- (factorial(5) * factorial(5)) / (factorial(9))
print(round(result_f_2, 5))
```

```
result_f_3 <- (factorial(5) * factorial(6)*
choose(200, 5)* choose(400, 5)) /
(factorial(10) *choose(200, 5)* choose(400, 5))
round(result_f_3, 6)
```

```
result_f_4 <- (factorial(5) * factorial(5)*
choose(200, 5)* choose(400, 5)) / (factorial(9)
*choose(200, 5)* choose(400, 5))
round(result_f_4, 6)
```

(d) Question 4:

R Code for Question 4b:

```
check_PMF_validity <- function(p1, p2, p3, p4, p5, p6) {
  p_values <- c(1 - p1, p1 * (1 - p2), p1 * p2 * (1 - p3),
p1 * p2 * p3 * (1 - p4), p1 * p2 * p3 * p4 * (1 - p5), p1
* p2 * p3 * p4 * p5 * (1 - p6), p1 * p2 * p3 * p4 * p5 * p6)
  total <- sum(p_values)
  is_valid <- abs(total - 1) < 1e-9
  return(list(sum = total, is_valid = is_valid))
}
```

# Test the function with random values

```
p1 <- runif(1)
p2 <- runif(1)
p3 <- runif(1)
p4 <- runif(1)
p5 <- runif(1)
p6 <- runif(1)
```

```
cat("p1:", p1, "\np2:", p2, "\np3:", p3, "\np4:",
p4, "\np5:", p5, "\np6:", p6, "\n")
```

```
result <- check_PMF_validity(p1, p2, p3, p4, p5, p6)
cat("Sum:", result[["sum"]], "\nIs Valid:", result[["is_valid"]], "\n")
```