Homework 1

Mengxiang Jiang

Stat 610 Distribution Theory

September 5, 2025

Problem 1. *Statistical Inference* by Casella and Berger, 2nd Edition, Chapter 1, Exercise 4, 5, and 6.

4 For events A and B, find formulas for the probabilities of the following events in terms of the quantities $P(A)$, $P(B)$, and $P(A \cap B)$.

(a) either A or B or both

(b) either A or B but not both

(c) at least one of A or B

(d) at most one of A or B

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

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

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

(d) $1 - P(A \cap B)$

5 Approximately one-third of all human twins are identical (one-egg) and two-thirds are fraternal (two-egg) twins. Identical twins are necessarily the same sex, with male and female being equally likely. Among fraternal twins, approximately one-fourth are both female, one-fourth are both male, and half are one male and one female. Finally, among all U.S. births, approximately 1 in 90 is a twin birth. Define the following events:

- A = the birth results in twin females
- B = the twins are identical twins
- C = a U.S. birth results in twins

(a) State, in words, the event $A \cap B \cap C$.

(b) Find $P(A \cap B \cap C)$.

(a) The event that a U.S. birth results in identical twin females.

(b) From the given information, we have

* $P(C) = \frac{1}{90}$
* $P(B|C) = \frac{1}{3}$
* $P(A|B \cap C) = \frac{1}{2}$

So, $P(A \cap B \cap C) = P(C)P(B|C)P(A|B \cap C) = \frac{1}{90} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{540}$.

6 Two pennies, one with $P(\text{head}) = u$ and one with $P(\text{head}) = w$, are to be tossed together independently. Define

- $p_0 = P(0 \text{ heads occur})$,
- $p_1 = P(1 \text{ head occurs})$,
- $p_2 = P(2 \text{ heads occur})$.

Can u and w be chosen so that $p_0 = p_1 = p_2$? Prove your answer.

We have

- $p_0 = (1-u)(1-w)$,
- $p_1 = u(1-w) + w(1-u)$,
- $p_2 = uw$.

So, $p_0 = p_2$ implies $(1-u)(1-w) = uw$, which simplifies to $u + w = 1$. Also, $p_1 = p_2$ implies $u(1-w) + w(1-u) = uw$, which simplifies to $u + w = 3uw$. Combining these two equations, we have $3uw = 1$, or $uw = \frac{1}{3}$. However, since $u + w = 1$, by AM-GM inequality, we have $\frac{u+w}{2} = \frac{1}{2} \geq \sqrt{uw}$, which when squared gives $\frac{1}{4} \geq uw$. This contradicts $uw = \frac{1}{3}$. Therefore, there are no such u and w, assuming they are both in $[0, 1]$.

Problem 2. Here is a sample space: $\mathcal{S} = \{a, b, c\}$.

(a) Explicitly provide the σ-algebra of all subsets.

(b) Suppose one may only observe whether the outcome is a or not. Explicitly provide the smallest relevant σ-algebra. Hint: what events may be obtained by complements, unions and intersections, starting only with a?

(a) The σ-algebra of all subsets is

$$\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

(b) Starting with $\{a\}$, we may obtain $\{b, c\}$ by taking its complement. Then, we may obtain $\{a, b, c\}$ by taking the union of $\{a\}$ and $\{b, c\}$. Finally, we may obtain $\emptyset$ by taking the complement of $\{a, b, c\}$. Thus the smallest relevant σ-algebra is

$$\{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\}.$$

Problem 3. For each of the following, be sure to explicitly provide the sample space and the event indicated.

(a) I hand out 4 pieces of candy at random, each to a child. There are three children and some will get more than others. Find the probability that each child gets at least one piece of candy. Find the probability that one child gets it all.

(b) The local Chevy dealer has 5 trucks to give away: two Silverados (S_1 and S_2), three Tahoes (T_1, T_2, and T_3). Suppose they randomly select a truck, give it to a lucky customer, and then randomly select another for a second lucky customer. Find the probability that x Silverados are selected, where $x = 0, 1, 2$.

(a) The sample space is
$$\mathcal{S} = \{(c_1, c_2, c_3, c_4) : c_i \in \{1, 2, 3\}\}.$$
The event that each child gets at least one piece of candy is
$$\begin{aligned} A = \{&(1,1,2,3),(1,1,3,2),(1,2,1,3),(1,2,2,3),\\ &(1,2,3,1),(1,2,3,2),(1,2,3,3),(1,3,1,2),\\ &(1,3,2,1),(1,3,2,2),(1,3,2,3),(1,3,3,2),\\ &(2,1,1,3),(2,1,2,3),(2,1,3,1),(2,1,3,2),\\ &(2,1,3,3),(2,2,1,3),(2,2,3,1),(2,3,1,1),\\ &(2,3,1,2),(2,3,1,3),(2,3,2,1),(2,3,3,1),\\ &(3,1,1,2),(3,1,2,1),(3,1,2,2),(3,1,2,3),\\ &(3,1,3,2),(3,2,1,1),(3,2,1,2),(3,2,1,3),\\ &(3,2,2,1),(3,2,3,1),(3,3,1,2),(3,3,2,1)\}. \end{aligned}$$
The event that one child gets it all is
$$B = \{(1,1,1,1),(2,2,2,2),(3,3,3,3)\}.$$
Since each outcome in the sample space is equally likely, we have
$$P(A) = \frac{36}{3^4} = \frac{4}{9}, \quad P(B) = \frac{3}{3^4} = \frac{1}{27}.$$

(b) The sample space is
$$\mathcal{S} = \{(x, y) : x, y \in \{S_1, S_2, T_1, T_2, T_3\}, x \neq y\}.$$
The event that $x = 0$ Silverados are selected is
$$A = \{(T_1,T_2),(T_1,T_3),(T_2,T_1),(T_2,T_3),(T_3,T_1),(T_3,T_2)\}.$$
The event that $x = 1$ Silverados are selected is
$$\begin{aligned} B = \{&(S_1,T_1),(S_1,T_2),(S_1,T_3),\\ &(S_2,T_1),(S_2,T_2),(S_2,T_3),\\ &(T_1,S_1),(T_1,S_2),(T_2,S_1),\\ &(T_2,S_2),(T_3,S_1),(T_3,S_2)\}. \end{aligned}$$

The event that $x = 2$ Silverados are selected is

$$C = \{(S_1, S_2), (S_2, S_1)\}.$$

Since each outcome in the sample space is equally likely, we have

$$P(A) = \frac{6}{20} = \frac{3}{10}, \quad P(B) = \frac{12}{20} = \frac{3}{5}, \quad P(C) = \frac{2}{20} = \frac{1}{10}.$$

Problem 4. *Statistical Inference* by Casella and Berger, 2nd Edition, Chapter 1, Exercise 10. Use the case $n = 2$ (Thm. 1.1.4.d in the book) and mathematical induction to solvee this.

10 Formulate and prove a version of DeMorgan's Laws that applies to a finite collection of sets $A_1, A_2, \ldots, A_n$.

For a finite collection of sets $A_1, A_2, \ldots, A_n$, we conjecture that the formulation is

$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c.$$

The other law is simply setting $B_i = A_i^c$ for each i, and taking the complement of both sides, giving us

$$\bigcap_{i=1}^{n} B_i^c = \left(\bigcup_{i=1}^{n} B_i\right)^c.$$

Since taking complements on both sides of an equation preserves equality, we will only attempt to prove the first formulation by mathematical induction.
Base case: when $n = 2$, we have

$$\left(\bigcap_{i=1}^{2} A_i\right)^c = \bigcup_{i=1}^{2} A_i^c.$$

Proof of base case: An element x is in the left hand side if and only if x is not in the intersection of A_1 and A_2. This means x is not in A_1 or x is not in A_2. This is equivalent to saying x is in the complement of A_1 or x is in the complement of A_2, which is equivalent to saying x is in the union of the complements of A_1 and A_2.

Inductive hypothesis: Assume that for some $k \geq 2$,

$$\left(\bigcap_{i=1}^{k} A_i\right)^c = \bigcup_{i=1}^{k} A_i^c.$$

Inductive step: We want to show that

$$\left(\bigcap_{i=1}^{k+1} A_i\right)^c = \bigcup_{i=1}^{k+1} A_i^c.$$

Proof of inductive step: First we can rewrite the left hand side as

$$\left(\bigcap_{i=1}^{k+1} A_i\right)^c = \left(A_{k+1} \cap \bigcap_{i=1}^{k} A_i\right)^c.$$

By the case $n = 2$ of DeMorgan's Laws, we have

$$\left(A_{k+1} \cap \bigcap_{i=1}^{k} A_i\right)^c = A_{k+1}^c \cup \left(\bigcap_{i=1}^{k} A_i\right)^c.$$

By the inductive hypothesis, we have

$$A_{k+1}^c \cup \left(\bigcap_{i=1}^{k} A_i\right)^c = A_{k+1}^c \cup \bigcup_{i=1}^{k} A_i^c.$$

Finally, we have

$$A_{k+1}^c \cup \bigcup_{i=1}^{k} A_i^c = \bigcup_{i=1}^{k+1} A_i^c.$$

Problem 5. The lifetime X of a cell phone battery, inspected under stress, satisfies $\mathrm{P}(X > t) = e^{-t/10}$ for all $t > 0$.

(a) Show that, if $0 \le a < b$ then $\mathrm{P}(a < X \le b) = \frac{1}{10}\int_a^b e^{-t/10}dt$.

(b) Is there a similar internal representation for $\mathrm{P}(a < X \le b) = \int_a^b g(t)dt$ valid even if $a < 0 < b$ Hint: use an indicator function – it should not depend on either a or b?

(c) Let $a \to b$ to find $\mathrm{P}(X = b)$.

(a) We have

$$\mathrm{P}(a < X \le b) = \mathrm{P}(X > a) - \mathrm{P}(X > b) = e^{-a/10} - e^{-b/10}.$$

Also, we have

$$\frac{1}{10}\int_a^b e^{-t/10}dt = \left[-e^{-t/10}\right]_a^b = e^{-a/10} - e^{-b/10}.$$

Thus, $\mathrm{P}(a < X \le b) = \frac{1}{10}\int_a^b e^{-t/10}dt$.

(b) We can define

$$g(t) = \begin{cases} \frac{1}{10}e^{-t/10}, & t > 0, \\ 0, & t \le 0. \end{cases}$$

Then, for any $a < b$, we have

$$\mathrm{P}(a < X \le b) = \int_a^b g(t)dt.$$

(c) Letting $a \to b$, we have

$$\mathrm{P}(X = b) = \mathrm{P}(b < X \leq b) = \int_b^b g(t)dt = 0.$$

Problem 6. Suppose a communications system is designed with K circuits so that if one is in use another can be chosen. Assume j circuits are in use and a circuit is chosen at random (without replacement) until a circuit not in use is selected.

(a) What is the chance that at least two selections are required?

(b) What is the chance that exactly two selections are required?

(c) What is the chance that exactly i selections are required, for $i = 1, 2, \ldots$?

(a) If the first selection is a circuit that is in use, then at least two selections are required, so the probability is just $\frac{j}{K}$.

(b) The first selection must be a circuit that is in use, which happens with probability $\frac{j}{K}$, and the second selection must be a circuit that is not in use, which happens with probability $\frac{K-j}{K-1}$, given that the first selection is a circuit that is in use. Thus, the probability that exactly two selections are required is

$$\frac{j}{K} \times \frac{K-j}{K-1} = \frac{j(K-j)}{K(K-1)}.$$

(c) The chance that exactly i selections are required is

$$\frac{j}{K} \times \frac{j-1}{K-1} \times \cdots \times \frac{j-(i-2)}{K-(i-2)} \times \frac{K-j}{K-(i-1)},$$

since the first $i-1$ selections must be circuits that are in use, and the i-th selection must be a circuit that is not in use. (Note that this still handles the case $i \geq j+2$ correctly, since that guarantees a 0 in the numerator of one of the fractions, and we know selecting more than j in use circuits is not possible without replacement.)

Problem 7. Quordle is a daily word game in which you have 9 tries to guess four 5-letter words. Suppose the four words are selected at random (without replacement) from a dictionary of 2309 5-letter words (which is what a similar game has).

(a) How many possible 4 word combinations are there?

(b) Find the number of ways that k daily games have no words in common, and use that to find the probability that at least one word is repeated. This is similar to the birthday problem discussed in class, except the sampling method is different.

(c) Find the probability that at least one word is repeated for $k = 2, 3, \ldots, 20$, using **R**. Note: `choose(n,i)` gives the binomial coefficient $\binom{n}{i}$ and `lchoose(n,i)` is the logarithm of that.

(d) What is the smallest number of daily games for which the probability that at least one word is repeated is at least 0.50? (Later we will see that this is the median number of games to be played until a word is repeated.)

(a) The number of possible 4 word combinations is

$$\binom{2309}{4} = \frac{2309!}{4! \times 2305!} = \frac{2309 \times 2308 \times 2307 \times 2306}{4 \times 3 \times 2 \times 1} = 1181286914501.$$

(b) The number of ways that k daily games have no words in common is

$$\begin{aligned}
&\binom{2309}{4} \times \binom{2305}{4} \times \cdots \times \binom{2309-4(k-1)}{4} \\
&= \frac{2309!}{4! \times \cancel{2305!}} \times \frac{\cancel{2305!}}{4! \times \cancel{2301!}} \times \cdots \times \frac{\cancel{(2309-4(k-1))!}}{4! \times (2309-4k)!} \\
&= \frac{2309!}{(4!)^k \times (2309-4k)!}.
\end{aligned}$$

(c) Code adapted from `birthday.r`:

```r
### quordle.r
### compute probability that no 2 games have same words in k games
### solution is 2309!/((4!)^k*(2309-4k)!*(2309 choose 4)^k)

### computing the probabilities
kgames = seq(1,20)  # vector of choices for k
probs = exp(lfactorial(2309)-kgames*lfactorial(4)-
lfactorial(2309-4*kgames)-kgames*lchoose(2309, 4))
probs_c = 1-probs
cbind(kgames,probs,probs_c)

```

Listing 1: quordle.r

The output is

```
      kgames     probs      probs_c
 [1,]      1 1.0000000 1.470823e-12
 [2,]      2 0.9930841 6.915904e-03
 [3,]      3 0.9793836 2.061637e-02
 [4,]      4 0.9591691 4.083088e-02
 [5,]      5 0.9328414 6.715864e-02
 [6,]      6 0.9009182 9.908179e-02
 [7,]      7 0.8640176 1.359824e-01
 [8,]      8 0.8228376 1.771624e-01
 [9,]      9 0.7781344 2.218656e-01
[10,]     10 0.7306994 2.693006e-01
[11,]     11 0.6813356 3.186644e-01
[12,]     12 0.6308357 3.691643e-01
[13,]     13 0.5799610 4.200390e-01
[14,]     14 0.5294236 4.705764e-01
```

```
[15,]       15  0.4798707  5.201293e-01
[16,]       16  0.4318731  5.681269e-01
[17,]       17  0.3859167  6.140833e-01
[18,]       18  0.3423978  6.576022e-01
[19,]       19  0.3016219  6.983781e-01
[20,]       20  0.2638053  7.361947e-01

```

Listing 2: Output of quordle.r

The probability that at least one word is repeated is listed in the `probs_c` column.

(d) As seen in the output above, the smallest number of daily games for which the probability that at least one word is repeated is at least 0.50 is 15.