Full Name: ______________________

Student No: ______________________

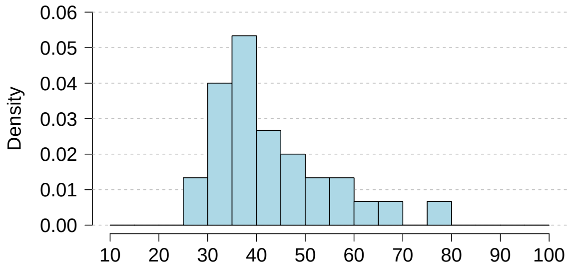
Grade:

Read before you start:

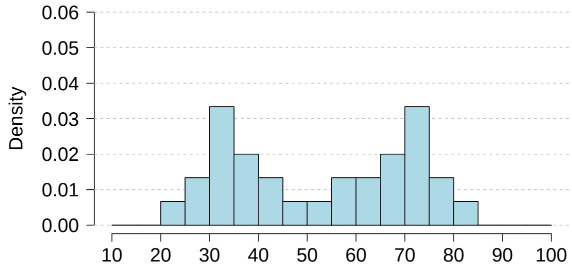
- Please make sure you write your full name and student number.
- The exam consists of 5 questions, each with multiple parts, and a total score of 80 points. All answers require justifications.
- The duration of the exam is 2 hours.

You can use the remainder of this page as scratch paper.

1. (15 points) The following two histograms illustrate the distribution of a numerical variable X within two categories of the same population.



category 1



category 2

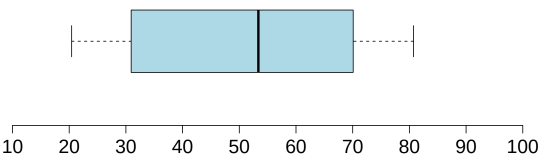
(a) Describe the qualitative shapes of the two histograms using terms such as *symmetric*, *right/left-skewed*, *unimodal*, etc.

Solution: The histogram of category 1 is unimodal and skewed to the right. The histogram of category 2 is bimodal and symmetric.

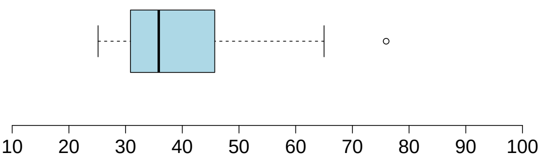
(b) The following are the boxplots of the distribution of X over the same two categories. Match the boxplots with the corresponding histograms.

Solution: The left boxplot corresponds to category 2, the right boxplot corresponds to category 1.

To see this, we can for instance notice that the histogram of category 1 shows no cases with values below 25, whereas in the left boxplot, the whisker for the smallest value is at (or close to) 20. Alternatively, the histogram of category 2 indicates some cases with values between 80 and 85, but the the largest value shown on the right boxplot is an outlier at below 80.



category ? $\rightarrow$ 2



category ? $\rightarrow$ 1

(c) Which of the two categories has a higher proportion of cases with values between 20 and 50?

Solution: Category 1. This can be seen either from the histograms or from the boxplots. The latter is easier.

From the histograms: We must compare the total area of the rectangles above the interval $[20, 50]$ in the two histograms. The height of the rectangles can only be determined approximately. However, it is evident that, the area of the rectangle above each of the intervals $[25, 30)$, $[30, 35)$, $[35, 40)$, $[40, 45)$ and $[45, 50)$ in the histogram of category 1 is at least as large as the area of the corresponding rectangle in the histogram of category 2. Furthermore, the area of the rectangle above $[35, 40)$ in category 1 is in fact larger than the sum of the areas of the rectangles above $[20, 25)$ and $[35, 40)$ in category 2. So, the total area above $[20, 50)$ in category 1 is larger than the corresponding are in category 2.

From the boxplots: From the boxplots, we can see that the median of X among category 1 is between 30 and 40, whereas the median of X among category 2 is larger than 50. This in particular means that more than 50% of the cases in category 1 have values between 20 and 50, and less than 50% of the cases in category 2 have values between 20 and 50.

2. (15 points) We have an ordinary 6-sided die.

(a) What is the probability that in 6 rolls of the die, we get 6 different numbers?

Solution: The outcome of the experiment can be described by the numbers seen on the die in the 6 rolls. The sample space is therefore $\Omega = \{1,2,3,4,5,6\}^6$. Since the die is fair, all the possible outcomes are equally likely, each having a probability of $1/|\Omega| = 1/6^6$. Let E denote the event that we get 6 different numbers. The number of outcomes realizing E is $6!$. (Namely, we have 6 possibilities for the first roll, 5 possibilities for the second roll, and so on.) Therefore,

$$\mathbb{P}(E) = |E| \cdot \frac{1}{|\Omega|} = \boxed{\frac{6!}{6^6}} \approx 0.0154321 \ .$$

(b) What is the probability that in 6 rolls of the die, we get at least 2 different numbers?

Solution: The sample space and the measure of probabilities are the same as the one in the previous part. Let F denote the event that we get at least 2 different numbers. The complement of this event is the event that we get the same number in all 6 rolls. That number can be any of the numbers $1, 2, \ldots, 6$, so there are exactly 6 possible outcomes realizing F^{c}. Therefore,

$$\mathbb{P}(F) = 1 - \mathbb{P}(F^{\mathsf{c}}) = 1 - \frac{|F^{\mathsf{c}}|}{|\Omega|} = 1 - \frac{6}{6^6} = \boxed{1 - \frac{1}{6^5}} \approx 0.9998714$$

(c) How many times do we have to roll the die so that, with probability at least $1/3$, we encounter at least one 5? (Find the minimum number of times with this property.)

Solution: We are looking for the smallest value n such that, in n rolls of the die, the probability of getting at least one 5 is larger than or equal to $1/3$.

Let A_n denote the event that, in n rolls of the die, we see at least one 5. This is an event in the sample space $\Omega_n = \{1,2,3,4,5,6\}^n$, which corresponds to the random experiment of rolling the die n times. Note that

$$\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^{\mathsf{c}}) = 1 - (5/6)^n \ .$$

In order for this probability to be at least $1/3$, we must have $(5/6)^n \leq 2/3$, or equivalently, $n \log(5/6) \leq \log(2/3)$, that is,

$$n \geq \frac{\log(2/3)}{\log(5/6)} \approx 2.223901 \ .$$

(Note that $\log(5/6)$ is negative, and that is why, in the last step, we have changed the direction of the inequality.) In conclusion, we need to roll the die at least $\boxed{3}$ times in order to make the probability of getting at least one 5 larger than or equal to $1/3$.

3. (25 points) Nine students are to be assigned to 3 different projects. To this end, the instructor asks each of the students to pick one of the numbers 1, 2, or 3 at random without communicating with the others. The students who pick number i (for $i = 1, 2, 3$) will then be assigned to work together on project number i.

(a) Choose an appropriate sample space for this random experiment and identify the probability of each possible outcome.

Solution: Number the students from 1 to 9. Each possible outcome can be described by a sequence of 9 numbers indicating the choices of the students. The sample space is thus the set

$$\Omega = \{(a_1, a_2, \dots, a_9) : a_1, a_2, \dots, a_9 \in \{1, 2, 3\}\} = \{1, 2, 3\}^9 .$$

of all such sequences. The choices of the students are independent of one another, thus the probability of each possible outcome $\omega = (a_1, a_2, \dots, a_9) \in \Omega$ is

$$\mathbb{P}(\omega) = \overbrace{1/3 \times 1/3 \times \cdots \times 1/3}^{\text{9 times}} = (1/3)^9 .$$

(b) What is the probability that project number 1 is assigned to at least one student?

Solution: Let A denote the event that project number 1 is chosen by at least one student. The complement of A is the event that none of the students has chosen project number 1. Therefore,

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = \boxed{1 - (2/3)^9} \approx 0.9739877 .$$

(c) What is the probability that each project is assigned to exactly 3 students?

Solution: Let B denote the event that each project is assigned to exactly 3 students. Since every possible outcome has the same probability $(1/3)^9$, we have $\mathbb{P}(B) = |B| \cdot (1/3)^9$. Hence, in order to find $\mathbb{P}(B)$, we need to count the number of elements in B.

The number of elements in B is $\frac{9!}{3!3!3!}$. One way to see this is as follows: In order to allocate the 9 students into 3 groups of equal size, first order them arbitrarily. Then, allocate the first 3 to group 1, the second 3 to group 2, and the last 3 to group 3.

$$\underbrace{\boxed{4}\ \boxed{1}\ \boxed{3}}_{\text{group 1}}\ \underbrace{\boxed{7}\ \boxed{9}\ \boxed{2}}_{\text{group 2}}\ \underbrace{\boxed{5}\ \boxed{6}\ \boxed{8}}_{\text{group 3}}$$

There are $9!$ different orderings. However, the ordering of the three students in each group is immaterial, so in the number $9!$, each distinct allocation is counted $3!\,3!\,3!$ times. Thus, dividing $9!$ by the number of repetitions gives the answer.

We conclude that

$$\mathbb{P}(B) = |B| \cdot \left(\frac{1}{3}\right)^9 = \frac{9!}{3!\,3!\,3!}\left(\frac{1}{3}\right)^9 = \boxed{\frac{9!}{(3!)^3 \cdot 3^9}} \approx 0.08535284 .$$

(d) What is the probability that two friends Jad and Laila end up in the same group?

Solution: Let C denote the event that Jad and Laila are end up in the same group. By the principle of total probability, we have

$$\begin{aligned}
\mathbb{P}(C) &= \sum_{i=1}^{3} \mathbb{P}(\text{Jad chooses project } i\text{, and } C \text{ occurs}) \\
&= \sum_{i=1}^{3} \mathbb{P}(\text{Jad chooses project } i)\, \mathbb{P}(C \mid \text{Jad chooses project } i) \\
&= \sum_{i=1}^{3} \mathbb{P}(\text{Jad chooses project } i)\, \mathbb{P}(\text{Laila chooses project } i) \\
&= \sum_{i=1}^{3} \left(\frac{1}{3}\right) \cdot \left(\frac{1}{3}\right) \\
&= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \boxed{\frac{1}{3}}\,.
\end{aligned}$$

Intuitively, whichever project Jad chooses, the chance of Laila choosing the same project is $1/3$.

(e) What is the conditional probability that Jad and Laila end up in the same group if we know that each project is assigned to exactly 3 students?

Solution: Let B and C be the events introduced in the previous parts, that is,

$$\begin{aligned}
B &= \langle \text{each project is assigned to exactly 3 students} \rangle\,, \\
C &= \langle \text{Jad and Laila end up in the same group} \rangle\,.
\end{aligned}$$

We are looking for

$$\mathbb{P}(C \mid B) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(B)}\,.$$

From part (c), we know $\mathbb{P}(B)$. It remains to find $\mathbb{P}(B \cap C)$. Since all the possible outcomes in the sample space are equally likely, we need to find $|B \cap C|$.
The set $B \cap C$ represents the event that every project has exactly 3 students assigned to it and Jad and Laila are assigned to the same project. We can count the number of elements in $|B \cap C|$ as follows: We have 3 possibilities for the project assigned to Jad and Laila. Then, there are 7 options for the third member of their group. Then, we are left with 6 students who must to be allocated to the remaining two project, with each project receiving exactly 3 students. The latter is similar to part (c) and can be done in $\frac{6!}{3!\,3!}$ different ways. Thus, in overall, we have

$$|B \cap C| = 3 \times 7 \times \frac{6!}{3!\,3!}\,,$$

which implies

$$\mathbb{P}(B \cap C) = 3 \times 7 \times \frac{6!}{3!\,3!} \left(\frac{1}{3}\right)^9$$

We conclude that

$$\mathbb{P}(C \mid B) = \frac{3 \times 7 \times \frac{6!}{3!\,3!} \left(\frac{1}{3}\right)^9}{\frac{9!}{3!\,3!\,3!} \left(\frac{1}{3}\right)^9} = \frac{3 \times 7 \times 6! \times 3!}{9!} = \boxed{1/4}\,.$$

4. (10 points) We have three boxes each containing 10 balls. Box #1 contains 9 blue balls and 1 red ball, box #2 contains 3 blue balls and 7 red balls, and box #3 contains 4 red balls, 6 yellow balls and no blue balls. We pick a box at random, and draw a ball from it at random.

   (a) What is the probability that the drawn ball is blue.

Solution: Let B stand for the event that the drawn ball is blue. We divide the possibilities based on which box was picked and calculate $\mathbb{P}(B)$ using the principle of total probability.
For $i = 1, 2, 3$, let A_i denote the event that box $\#i$ is chosen. We have

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1) + \mathbb{P}(A_2)\,\mathbb{P}(B \mid A_2) + \mathbb{P}(A_3)\,\mathbb{P}(B \mid A_3) \\ &= \frac{1}{3} \times \frac{9}{10} + \frac{1}{3} \times \frac{3}{10} + \frac{1}{3} \times \frac{0}{10} \\ &= \boxed{2/5}\,.\end{aligned}$$

   (b) Assuming that the drawn ball is blue, what is the probability that it comes from box #1?

Solution: With the notation introduced above, we are looking for $\mathbb{P}(A_1 \mid B)$. We have

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1)}{\mathbb{P}(B)} = \frac{(1/3) \times (9/10)}{2/5} = \boxed{3/4}\,,$$

where for the denominator, we have used the result of the previous part.

5. (15 points) A jar contains 5 blue balls, 7 red balls, and 8 yellow balls. Jad and Laila play the following game: They draw three balls from the jar at random without replacement. If the drawn balls all have the same colour, Jad wins, and if they all have different colours, Laila wins. If neither of these two possibilities occurs, then the game is a draw (i.e., neither player wins).

   (a) Is this game fair? If not, which of the two players has an advantage over the other?

Solution: The game would be fair if Jad and Laila have the same probability of winning. Let us compute the two probabilities. Let J denote the event that Jad wins and L the event that Laila wins.
The sample space Ω is the collection of all 3-element subsets of the 20 balls in the jar. All the possible outcomes are equally likely, so in order to find the probabilities of J and L, we need to find the number of elements in J, L, and Ω. We have

$$\begin{aligned}|\Omega| &= \binom{20}{3} = 1140\,, \\ |J| &= \binom{5}{3} + \binom{7}{3} + \binom{8}{3} = 10 + 35 + 56 = 101\,, \\ |L| &= 5 \times 7 \times 8 = 280\,.\end{aligned}$$

(Too count the number of elements in J, we note that there are $\binom{5}{3}$ possibilities to choose 3 blue balls, $\binom{7}{3}$ possibilities to choose 3 red balls, and $\binom{8}{3}$ possibilities to choose 3 yellow balls.)
Therefore,

$$\mathbb{P}(J) = \frac{|J|}{|\Omega|} = \boxed{\frac{101}{1140}} \approx 0.08859649\,, \qquad \mathbb{P}(L) = \frac{|L|}{|\Omega|} = \boxed{\frac{280}{1140}} \approx 0.245614\,.$$

We conclude that the game is unfair, and that Laila has an advantage over Jad.

(b) What is the chance that the three drawn balls are blue if we learn that Jad has won?

Solution: Let B denote the event that 3 blue balls are drawn. We have

$$\mathbb{P}(B \mid J) = \frac{\mathbb{P}(B \cap J)}{|J|} = \frac{|B\cap J|/|\Omega|}{|J|/|\Omega|} = \frac{|B \cap J|}{|J|} = \frac{\binom{5}{3}}{\binom{5}{3} + \binom{7}{3} + \binom{8}{3}} = \boxed{\frac{10}{101}} \approx 0.0990099 \ .$$

(c) Suppose the players repeat playing until one of them wins for the first time. What is the chance that the eventual winner is Jad?

Solution: We start with a intuitive answer and then provide a more rigorous reasoning.

Intuitive reasoning: It is clear that, if the game is repeated over and over, one of the players will eventually win. Consider the first time in which one of the players wins. For that game, we know that the event $J \cup L$ has happened, and we are looking for the conditional probability that J has happened. Since the previous games have no way of affecting this last game, the probability we are looking for is nothing but

$$\mathbb{P}(J \mid J \cup L) = \frac{\mathbb{P}(J)}{\mathbb{P}(J \cup L)} = \frac{101/1140}{101/1140 + 280/1140} = \boxed{\frac{101}{381}} \approx 0.2650919 \ .$$

Rigorous reasoning: We have a new "super-experiment" consisting of repetitions of the original game. We can describe each possible outcome of the super-experiment by a string of the form $\texttt{DD}\cdots\texttt{DJ}$ (if Jad eventually wins) or $\texttt{DD}\cdots\texttt{DL}$ (if Laila eventually wins). The $\texttt{D}$'s in front correspond to the games that ended in a draw. The sample space is the set of all such strings, that is,

$$\Omega^* = \{\texttt{D}^n\texttt{J}, \texttt{D}^n\texttt{L} : n = 0, 1, 2, \ldots\} \ .$$

From part (a), we know that the probability that Jad wins in one game is $p_J := 101/1140$, and the probability that Laila wins in one game is $p_L := 280/1140$. Let $J^* \subseteq \Omega^*$ denote the event that Jad eventually wins, and for each $k = 1, 2, 3, \ldots$, let A_k denote the event that the first time in which one of the players wins is in game number k. Note that $A_1, A_2, \ldots$ partition the sample space. Hence, by the principle of total probability,

$$\mathbb{P}(J^*) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)\, \mathbb{P}(J^* \mid A_k) \ .$$

But for every k,

$$\mathbb{P}(J^* \mid A_k) = \frac{p_J}{p_J + p_L} \ .$$

Therefore,

$$\mathbb{P}(J^*) = \sum_{k=1}^{\infty} \mathbb{P}(A_k) \cdot \frac{p_J}{p_J + p_L} = \Bigg(\underbrace{\sum_{k=1}^{\infty} \mathbb{P}(A_k)}_{=1}\Bigg) \frac{p_J}{p_J + p_L}$$

$$= \frac{p_J}{p_J + p_L} = \frac{101/1140}{101/1140 + 280/1140} = \boxed{\frac{101}{381}} \approx 0.2650919 \ .$$