CIT 5920–Mathematical Foundations of Computer Science
Homework 4: *Introduction to discrete probability*

Version: October 10, 2024

Complete the online portion on PrairieLearn. For the written section, use the HW1 template on Overleaf, shared on the course forum and Canvas. Each exercise should be on a separate page. Only submissions in this format will be accepted. Submit your written work on Gradescope by the deadline. For assistance or inquiries, don't hesitate to: Attend office hours; post questions on the class forum; ask about the motivation behind this material.

Guidelines:

- Ensure clarity in your answers.
- Avoid direct answers unless specified. Merely providing a number will result in deductions.
- Clearly state any assumptions in combinatorics questions. Reasonable assumptions will be credited. For instance, assuming two people are identical is not reasonable, neither is assuming that the order matters when mixing paints.
- Prefer math symbols over words, e.g., use $A \cap B$ instead of "The elements common to both A and B."
- If unsure about the length or style of your answer, consult during office hours or post on the class forum. However, we won't provide direct answers.
- For the PrairieLearn section, consider using Wolfram Alpha's Equation Solver. For the written section, avoid calculators. Answers can be in factorial or $\binom{n}{k}$ form.

Exercise 0 – PrairieLearn Questions [14pts]

Use the QR code to the right to access the PrairieLearn portion of this homework. **Please login using your Penn Google account.**

1. Most questions are designed to provide you with an infinite number of variations.
2. With these questions, if you respond incorrectly, you will have the opportunity to try again until you get the question right. To earn credit on the question, you must answer *any* variant from the first try.

Exercise 1 – Probability of Specific Card Hands [4pts]

Recall that a deck of cards has 52 cards across 4 suits, 12 of which are face cards, equally split between Kings, queens and jacks.

A 5-card hand is dealt from a perfectly shuffled deck of playing cards. Determine the probability of each of the following events:

A. The hand has at least one card of the suit of clubs.

B. The hand has at least two cards with the same rank (for instance, two cards that are Kings are of the same rank, as are two cards that are 3's).

C. The hand has exactly one club or exactly one spade.

D. The hand has at least one club or at least one spade.

Exercise 2 – Relating Discrete Math Concepts to Cultural and Familiar Examples [10pts]

In this course, we often use common examples like coins, dice, and playing cards to make abstract concepts in discrete mathematics more tangible and relatable. These examples serve as a bridge between the theoretical and the practical, helping students understand and visualize complex ideas.

> **EXAMPLE 7.3 (Deck of cards)** A card is drawn from an ordinary deck of 52 cards which is pictured in Fig. 7-2(a).
>
> The sample space S consists of the four *suits*, clubs (C), diamonds (D), hearts (H), and spades (S), where each suit contains 13 cards which are numbered 2 to 10, and jack (J), queen (Q), king (K), and ace (A). The hearts (H) and diamonds (D) are red cards, and the spades (S) and clubs (C) are black cards. Figure 7-2(b) pictures 52 points which represent the deck S of cards in the obvious way. Let E be the event of a *picture card, or face card*, that is, a Jack (J), Queen (Q), or King (K), and let F be the event of a heart. Then $E \cap F = \{JH, QH, KH\}$, as shaded in Fig. 7-2($b$).

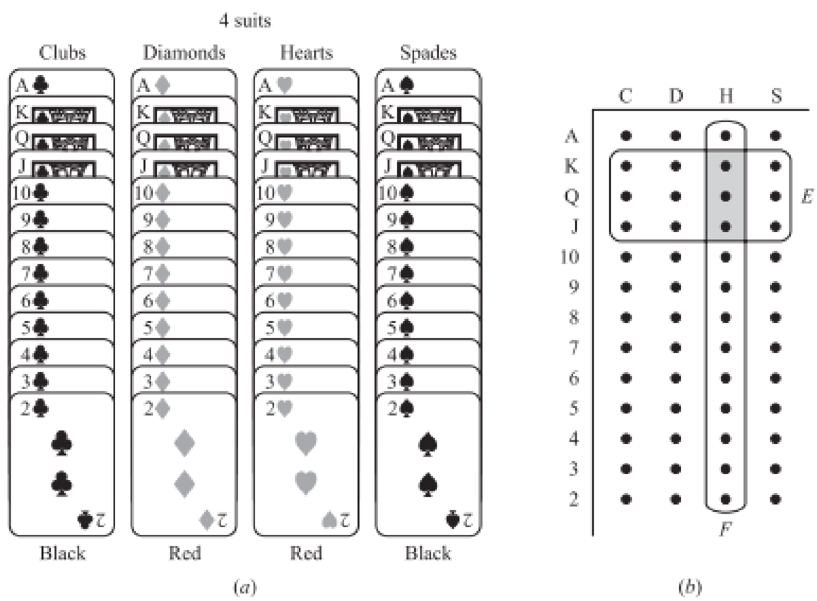

Fig. 7-2

Figure 1: Illustration of all playing cards, organized in four columns by color. Face cards are outlined, and counts are provided for each category. From Schaum's Outlines of Discrete Mathematics.

Refer to Figure 1 which provides a visual representation of all playing cards. This illustration is an example of how we can visually represent a familiar concept to aid in understanding.

However, it's essential to recognize that what's familiar to one person might not be to another. The examples we use, while practical, might not resonate with everyone due to cultural or personal differences. Finding universally relatable examples is challenging, but the diversity of our student body can be a strength in this endeavor.

Your Task:

1. Reflect on your cultural background, personal experiences, and favorite board or video games.
2. Identify an example from your reflection that can be used to illustrate counting or probability concepts, similar to how coins, dice, and playing cards are used in this course.

3. Describe the features of your chosen example, the space of outcomes, and any other relevant details. *What is most important is that you make it very clear are the different combinatorial elements, the different choices, see examples for more details.*

4. If possible, provide an illustration or visual representation of your example. While LaTeX-produced PDFs are preferred, for this exercise, you may also submit vector art produced by a tablet or other tools.

Here are some examples to inspire you, the important part is the combinatorial elements:

- **Pachisi:** An ancient Indian board game where players move their pawns based on the throw of six or seven cowrie shells, with the number of shells landing with the opening upwards determining the number of spaces the player moves.
  - *Combinatorial Elements:* Players utilize six or seven cowrie shells for determining moves. Each shell has two distinct sides: opening upwards and opening downwards. The combination of how these shells land (e.g., 5 shells with opening upwards and 2 with opening downwards) dictates the number of spaces moved.
- **War:** A simple card game where the deck is divided evenly among two players, and each player reveals the top card of their deck. The player with the higher card wins both cards.
  - *Combinatorial Elements:* The game uses a standard deck of 52 cards. Each card has a rank ranging from 1 to 13. The suits and colors are not relevant for the outcome of the game, as only the rank determines the winner of each round.
- **UNO:** A popular card game where players match cards by color or number, use action cards to mix up the game, and aim to be the first to get rid of all their cards.
  - *Combinatorial Elements:* The game consists of 108 cards. These are divided into four colors: red, yellow, green, and blue. Each color has cards numbered from 0 to 9, and special action cards (Skip, Reverse, and Draw Two). Additionally, there are Wild and Wild Draw Four cards that are colorless.

Submissions will be graded based on thoughtfulness, the likelihood that these examples might be useful to students, and creativity/originality. Any submission deemed good enough to be used in future courses will receive extra credit.

Remember, the goal is to find examples that make abstract concepts feel real and relatable to one of your peers.

Exercise 3 – Probability on Permutations with Manual Combinations [6pts]

The digits 1, 2, 3, and 4 are randomly arranged to form two two-digit numbers. Each digit can only be used once. For example, if one of the two-digit numbers is 42, then the other two-digit number is either 13 or 31. What is the expected value of the product of the two numbers?

Exercise 4 – Probability on Permutations with Indicator Variables [10pts]

A family with a mother, father, two daughters, and three sons lines up in a random order for a photo.

A. Consider a scenario where you select any person in the family other than the mother. How many ways can you arrange the family such that the selected person ends up standing next to the mother?

B. Given the family has 7 members in total, and considering the previous arrangement, what is the probability that the selected person is next to the mother in a random arrangement?

C. Let's introduce the concept of indicator variables. An indicator variable takes the value 1 if a particular condition is met and 0 otherwise. Define D_1 and D_2 as indicator variables for the two daughters such that $D_i = 1$ if daughter i is standing next to the mother and $D_i = 0$ otherwise. How can you represent the total number of daughters standing next to the mother using D_1 and D_2?

D. Using the previously calculated probability and the indicator variables D_1 and D_2, what is the expected number of daughters standing next to the mother, $\mathbb{E}[D]$?

E. Similarly, let N_i be indicator variables for the sons. Using a similar approach, what is the expected number of sons standing next to the mother, $\mathbb{E}[N]$?

Exercise 5 – Email Spam Filter [6pts]

An email service uses a spam filter that flags emails containing the word "offer" as potential spam. The following statistics are known:

- 5% of all emails are spam.
- 70% of spam emails contain the word "offer".
- 10% of legitimate emails contain the word "offer".

We are interested in finding out: If an email contains the word "offer", what is the probability that it is spam?

A. Define the events S and O. What are the values of $P(S)$ and $P(\overline{S})$?

B. What are the values of $P(O|S)$ and $P(O|\overline{S})$?

C. Use the Law of Total Probability to calculate $P(O)$.

D. Apply Bayes' Theorem to find $P(S|O)$.

E. Interpret the result. What is the probability that an email containing "offer" is spam?

Exercise 6 – Machine Learning Classifier Performance [6pts]

A machine learning model classifies images as "cat" or "not cat". The following data is known:

- 20% of all images are actually of cats.
- The model correctly identifies cat images 95% of the time.
- It incorrectly labels non-cat images as "cat" 10% of the time.

If the model labels an image as "cat", what is the probability that it actually is a cat?

A. Define the events C and L. What are the values of $P(C)$ and $P(\overline{C})$?

B. What are the values of $P(L|C)$ and $P(L|\overline{C})$, where L represents the model labeling the image as "cat"?

C. Calculate $P(L)$ using the Law of Total Probability.

D. Use Bayes' Theorem to compute $P(C|L)$.

E. What does this result mean in the context of the model's performance?

Exercise 7 – Financial Risk Assessment [6pts]

A bank assesses loan applicants for default risk:

- 2% of applicants are high risk (will default).

- The risk assessment tool correctly identifies high-risk applicants 95% of the time.
- It incorrectly labels low-risk applicants as high risk 5% of the time.

If an applicant is identified as high risk, what is the probability they will default?

A. Define the events D and H. What are the values of $P(D)$ and $P(\overline{D})$?

B. What are the values of $P(H|D)$ and $P(H|\overline{D})$, where H represents the applicant being identified as high risk?

C. Calculate $P(H)$ using the Law of Total Probability.

D. Use Bayes' Theorem to find $P(D|H)$.

E. Interpret this probability. What does it say about the risk assessment tool's prediction?

Exercise 8 – Hash Table Occupancy and Collisions [6pts]

A **hash table** is a data structure used to store and retrieve data efficiently. It uses a **hash function** to map **keys** (which could be numbers, words, or other data) to **slots** in an array. The goal is for the hash function to distribute the keys uniformly across the slots to minimize the chance that two keys map to the same slot, which is called a **collision**.

In this exercise, we'll explore how probability helps us understand the behavior of hash tables, especially regarding empty slots and collisions.

Suppose you have a hash table with n slots, and you insert k keys into it. Each key is hashed independently and uniformly at random into one of the n slots. This means every key has an equal chance of going into any slot, and the placement of one key doesn't affect the placement of another.

A. Let X be the number of empty slots after all k keys have been inserted into the hash table. Express the expected number of empty slots, $\mathbb{E}[X]$, in terms of n and k.

B. Suppose the hash table has $n = 1000$ slots and you insert $k = 1000$ keys into it. Compute the expected number of empty slots $\mathbb{E}[X]$.

C. For large n and when $k = n$, derive an approximation for $\mathbb{E}[X]$ using the natural exponential function e.

D. Explain why, even when the number of keys equals the number of slots ($k = n$), we expect some slots to remain empty. Discuss how this relates to the concepts of collisions and load factor in hash tables.