# Homework 10

*Due: Friday, April 25, 2025*

All homeworks are due at 11:59 PM on Gradescope.

**Please do not include any identifying information about yourself in the handin, including your Banner ID.**

Be sure to fully explain your reasoning and show all work for full credit.

Problems marked with a * are problems which may appear on the midterm or final with some modification.

## Problem 1

The CS22 HTAs have just finished a grading meeting, and meet up at the CS22 Saloon for some games of dice with Rob and Ellis.

For the 7 of them, Tyler brings along a fair die with 7 faces numbered 1 through 7.

The group gathers around for a simple game.

To start the game, each member picks a number from the hat, 1-7 (everyone will have a different number).

On each round of the game, the die is rolled and whoever has the number gets 1 point. The game ends after 10 rounds, and the winner is declared to be the person who has accumulated the most points.

For the first game, Christina's number is 3. *

a. Define a sample space and a random variable to represent the outcome of a single die roll.

b. What is the expected value of the number that comes up when the die is rolled once? Show your work using the random variable you defined in part 1.

c. What is the variance of the number that comes up when the die is rolled once? Show your work using the random variable you defined in part 1.

d. What is the probability that Christina's number is rolled 10 times in a row, and she wins without any one else getting a point?

e. What is the probability that Christina's number is rolled at least 3 times?

Now for game 2, Christina picks the number 6, and just her luck, as she had brought along a bag of 5 seven-sided dice, where 4 of them are standard, and one of them is a special weighted die that will only ever roll a 6. She convinces Tyler it will be more fun to use this bag of dice she "forgot about" until just now. The group will pick a die uniformly at random from this bag and play the game as before.

f. What is the expected value of the number that comes up when one of the dice from Christina's bag is rolled once?

g. After the game ends, the TAs are ready to go home, but Christina insists they keep rolling until her number is rolled one final time. Tyler chooses a die from her bag to use at random. Assuming no limit on rolls, what's the expected number of rolls it will take for Christina's number (6) to be rolled?

**Solution:**

a. The sample space is $S = \{1, \ldots, 7\}$ and the random variable $X(\omega) = \omega$ where $\omega \in S$.

b. To find $\mathbb{E}(X)$ we use the definition of expected value.
$\mathbb{E}(X) = 1(1/7) + 2(1/7) + 3(1/7) + 4(1/7) + 5(1/7) + 6(1/7) + 7(1/7)$
$\mathbb{E}(X) = 4$

c. We use the definition of variance to solve for $V(X)$:
$V(X) = (1-4)^2(1/7) + (2-4)^2)1/7... + (7-4)^2(1/7)$
$V(X) = 4$

d. The probability that $X = 3$ ten times in a row is $\Pr(X = 3)^2$ or $(1/7)^{10}$.

e. To find the probability that $X = 3$ at least 3 times, we subtract the probabilities of the events where it gets rolled 0 times, once, and twice:

$$1 - \binom{10}{2}(\frac{1}{7})^2(\frac{6}{7})^8 - \binom{10}{1}(\frac{1}{7})^1(\frac{6}{7})^9 - \binom{10}{0}(\frac{1}{7})^0(\frac{6}{7})^{10}$$

f. $\mathbb{E}(Y) = 1((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 0)) + 2((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 0)) + 3((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 0)) + 4((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 0)) + 5((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 0)) + 6((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 1)) + 7((\frac{4}{5}(\frac{1}{7}) + (\frac{1}{5} * 0))$

g. We use a mean time to failure analysis, in which the expected number of times it will take until $Y = n$ is $\frac{1}{\Pr(Y=n)}$.
Since we are rolling one die until we get 6, there are two cases: we choose a normal die or we choose the special weighted die.
Let Z be a random variable that maps the type of die (normal or special) to the number of times it needs to be rolled to obtain a 6.

- In case we roll a normal die, the probability of rolling a 6 is $\frac{1}{7}$. This means we expect to roll the die 7 times to get a 6.
- In case we roll the special weighted die, we will always roll a 6, so the probability will be 1. This means that we expect to roll the die only once to get a 6.

We have a $\frac{4}{5}$ chance of choosing a normal die and $\frac{1}{5}$ chance of choosing the special die, so $\Pr(Z = 7) = \frac{4}{5}$ and $\Pr(Z = 1) = \frac{1}{5}$.

Finally, $\mathbb{E}(Z) = \frac{4}{5} \times 7 + \frac{1}{5} \times 1 = \frac{29}{5} = 5.8$

So we expect it to take $\frac{29}{5}$ die rolls, or 5.8.

# Problem 2

You are going for a treasure hunt and reach a fork in the road where you can go either to the left or the right. Down one path (either the left or the right) is the treasure you seek and down the other is certain doom.

Luckily you have a handy device that can, with reasonable probability, tell you which path to go down. In particular, your device tells you to go either left or right. If the correct answer is left, it always tells you to go left. On the other hand, if the correct answer is right, it only tells you to go left $\frac{1}{10}$ of the time. You also know that treasure is almost always on the right side of a fork in the road: of all forks in the roads, treasure is on the right $\frac{7}{10}$ of the time.

Given that your device tells you to go left, what is the probability that the treasure is on the left?*

**Solution:**

We use Bayes' rule. Let $L$ and $R$ be the event of the treasure on the left and right respectively. Let $D$ be the event that the device tells us to go left.

By Bayes' rule we have

$$\Pr(L|D) = \Pr(L) \cdot \frac{\Pr(D|L)}{\Pr(D)} = .3 \cdot \frac{1}{\Pr(D)} \tag{1}$$

and

$$\Pr(R|D) = \Pr(R) \cdot \frac{\Pr(D|R)}{\Pr(D)} = .7 \cdot \frac{1/10}{\Pr(D)}$$

Also, notice that $L$ and $R$ partition our probability space so

$$.3 \cdot \frac{1}{\Pr(D)} + .7 \cdot \frac{1/10}{\Pr(D)} = \Pr(L|D) + \Pr(R|D) = 1$$

Solving for $\Pr(D)$ we get

$$\Pr(D) = .37$$

Plugging this into Equation 1, we get $\Pr(L|D) = 30/37$.

# Problem 3

Suppose you have two coins. One coin is a fair coin (there is a 50% chance if you flip it you get heads and a 50% chance if you flip it you get tails). The other coin has heads on both sides (so there is a 100% chance if you flip it you get heads). Your friend chooses one of these coins uniformly at random and then you flip it 12 times and get all heads. What is the probability that your friend chose the fair coin?*

**Solution:**

We use Bayes' rule. Let $F$ be the event your friend chooses the fair coin, $U$ be the event they choose the unfair coin (with heads on both sides) and let $H$ be the event of getting 12 heads. By Bayes' rule we have

$$\Pr(F|H) = \Pr(F) \cdot \frac{\Pr(H|F)}{\Pr(H)} = \tag{2}$$

$\Pr(F)$ is .5 since our friend chose a uniformly random coin. $\Pr(H|F)$ is $1/2^{12}$ since the probability of getting 12 heads in a row with a fair coin is $1/2^{12}$. Furthermore, since $F$ and $U$ partition our sample space we have

$$H = (F \cap H) \cup (U \cap H)$$

$(F \cap H)$ and $(U \cap H)$ are disjoint and so we know

$$\begin{aligned} \Pr(H) &= \Pr(F \cap H) + \Pr(U \cap H) \\ &= \Pr(H|F) \cdot \Pr(F) + \Pr(H|U) \cdot \Pr(U) \\ &= 1/2^{12} \cdot .5 + 1 \cdot .5 \end{aligned}$$

Plugging all this into Equation 2, we get

$$\Pr(F|H) = .5 \cdot \frac{1/2^{12}}{1/2^{13} + .5} = \frac{1}{4097}$$

# Problem 4

Way back in the day, we talked about *conjunctive normal form* (CNF) for propositional formulas. We define a *literal* to be either a propositional letter (for example: $p$, $q$, $\ldots$) or the negation of a propositional letter (for example: $\neg p$, $\neg r$, $\ldots$). A *clause* is a sequence of literals joined by $\vee$, with no propositional letter appearing more than once. A formula is in CNF if it is a sequence of clauses joined by $\wedge$: for example, $(p \vee \neg q \vee r) \wedge (\neg p \vee \neg s) \wedge (q \vee s)$.

In this problem we'll expand on this idea. A $k$-clause is a clause with exactly $k$ literals. Let $C = (c_1, \ldots, c_n)$ be a sequence of $n > 0$ distinct $k$-clauses. Let $V$ be the set of propositional letters that appear in the clauses in $C$. The clauses do not necessarily contain the same propositional letters.

> For example: let $k = 3$, $n = 4$, and $C = (p \vee \neg q \vee r,\ p \vee \neg s \vee t,\ \neg p \vee s \vee t,\ p \vee r \vee \neg s)$. Then $V = \{p, q, r, s, t\}$.

We will randomly assign true/false values to the propositional letters in $V$, with each value equally likely for each variable.

a. In terms of $k$ and $n$, what is the *smallest* possible value for $|V|$? What is the *largest* possible value? Justify your answers.

   (Note: you may assume that $n \leq k! \cdot 2^k$.)

b. Under the random assignment of truth values to letters, what is the probability that $c_n$ is true?

c. What is the expected number of true $k$-clauses in $C$?

d. If we connect the clauses in $C$ together with $\wedge$s, we get a formula in CNF. Using your answer to part c., prove that this CNF formula must be satisfiable when $n < 2^k$.

> **HINT:** A random variable cannot always be less than its expectation—why not?

**Solution:**

a. $|V|$ is minimized if all clauses in $C$ contain the same letters, and each one contains $k$ letters, so $|V| \geq k$. $|V|$ is maximized if all $n$ clauses contain different letters, so $|V| \leq nk$.

b. $c_n$ is true if at least one of its $k$ literals is true. The probability of any literal being *false* is $\frac{1}{2}$ and is independent of the probabilities of any of the other literals being false. So we can use the product rule to compute that the probability of all $k$ literals being false is $\frac{1}{2^k}$. The event we're interested in, at least one literal being true, is the complement of this event, and so its probability is $1 - \frac{1}{2^k}$.

c. Let $I_i$ be the indicator random variable that returns 1 if $c_i$ is true and 0 otherwise. We want to compute $\mathbb{E}[\sum_{i=1}^{n} I_i]$. By linearity of expectation this is $\sum_{i=1}^{n} \mathbb{E}[I_i]$, and by the previous part, $\mathbb{E}[I_i] = 1 - \frac{1}{2^k}$ for each $i$. So we expect $n(1 - \frac{1}{2^k})$ true clauses.

d. We first prove the lemma suggested in the hint: for any random variable $R$ over a probability space $\Omega$, there is some $\omega \in \Omega$ with $R(\omega) \geq \mathbb{E}[R]$. Suppose otherwise; then

$$\begin{aligned}
\mathbb{E}[R] &= \sum_{\omega \in \Omega} R(\omega)\Pr[\omega] \\
&< \sum_{\omega \in \Omega} \mathbb{E}[R]\Pr[\omega] \\
&= \mathbb{E}[R] \sum_{\omega \in \Omega} \Pr[\omega] \\
&= \mathbb{E}[R]
\end{aligned}$$

which cannot be.

Now suppose $n < 2^k$ and let $R = \sum_{i=1}^{n} I_i$. By our lemma, there is some truth assignment $\omega$ that makes at least $\mathbb{E}(R)$ clauses true. If we can show that $\mathbb{E}(R) > n - 1$, we are done: then $\omega$ must make at least $n$ clauses (i.e. all of them) true, and so $\omega$ is a satisfying assignment.

So our goal becomes to show that $\mathbb{E}(R) > n - 1$. By part c. we have that $\mathbb{E}[R] = n(1 - \frac{1}{2^k})$. Since $0 < n < 2^k$, we have that $\frac{1}{n} > \frac{1}{2^k}$, and thus $1 - \frac{1}{n} < 1 - \frac{1}{2^k}$. So

$$\begin{aligned}
\mathbb{E}[R] &= n(1 - \frac{1}{2^k}) \\
&> n(1 - \frac{1}{n}) \\
&= n - 1
\end{aligned}$$

as desired.

# Problem 5 (Mind Bender — *Extra Credit*)

Suppose we are given $n$ horses and $n$ unicorns. Each horse has a unique top speed that is an integer which is unknown to us. Unicorns are able to run at any speed they like. Next, suppose we arrange the horses and unicorns in a line (all facing the same direction) in a *uniformly random order* and let them loose to run.

Each horse then runs either its top speed, or, if the horse/unicorn in front of it in the line is going slower than its top speed, it matches the speed of the horse/unicorn in front of it. If a unicorn is behind another horse/unicorn that is going speed $s$, then it goes speed $s - 1/(3n)$. If a unicorn is at the front of the line, it goes strictly faster than all other horses and unicorns.

The result of this process is a series of "clumps" of horses/unicorns: in each clump every horse/unicorn is running the same speed and, in particular, the speed of the horse/unicorn at the front of the clump. Notice that the clumps towards the front of the line go faster than the clumps towards the back of the line.

Let $X$ be the random variable that gives the number of clumps under our uniformly random order. What is $\mathbb{E}[X]$ as a function of $n$? You may give an answer with $\sum$s in it.

**HINT:** One way of constructing a uniformly random permutation on set $A$ is to start with an arbitrary element $a \in A$, then take another arbitrary element $b \in A$ and place it to the left of $a$ with probability $1/2$ and to the right of $a$ with probability $1/2$ then take another arbitrary element $c \in A$ and place it in one of the three possible slots between to the left and right of $a$ and $b$, each with probability $1/3$ etc.

**Solution:**

As per the hint, we imagine that we construct our uniformly random permutation by first inserting all horses in ascending order of their speeds into our permutation then inserting all unicorns.

Notice that, when we insert a horse, since we are inserting our horses in ascending order of speed, the only way we create a new clump is if we insert our new horse to the front of the line. If we have inserted $i-1$ horses so far, the probability of this is $1/(i)$. As such, if we let $H_i$ be the indicator of whether the $i$th horse inserted creates a new clump, then we have $\mathbb{E}[H_i] = \frac{1}{i}$. It follows by linearity of expectation that the total number of clumps in expectation after we insert all horses is

$$\sum_{i=1}^{n} \mathbb{E}[H_i] = \sum_{i=1}^{n} \frac{1}{i}.$$

Next, notice that each time we insert a unicorn, we create a new clump with probability 1. As such, if $U_i$ is the indicator of whether inserting the $i$th unicorn creates a clump, we have

$$\sum_{i=1}^{n} \mathbb{E}[U_i] = n.$$

Lastly, by linearity of expectation we have

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[H_i] + \sum_{i=1}^{n} \mathbb{E}[U_i] = n = n + \sum_{i=1}^{n} \frac{1}{i}.$$