# Assignment 2

## Marked Questions

### Q1. [10 Points] Counting Labeled Trees *(maximum 2 pages)*

In this question we will walk you through the proof of the following claim.

**Claim 1.** *There are $n^{n-2}$ different labeled trees on $n$ vertices.*

a. (5 Points.) Label the vertices from 1 to $n$ and perform the following procedure to construct two lists $A$ and $B$:

   i. Take the leaf with the smallest label and its neighbour and add it to $B$ and $A$ respectively.
   ii. Remove this leaf.
   iii. Repeat until only one edge remains.

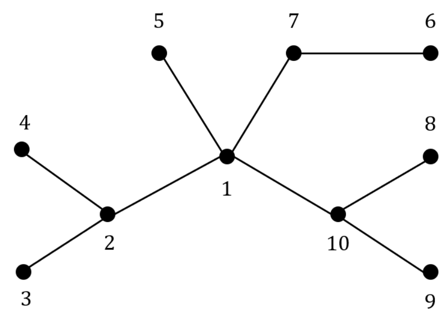



Figure 1: Example tree.

For the graph in Figure 1, $A = [2, 2, 1, 1, 7, 1, 10, 10]$ and $B = [3, 4, 2, 5, 6, 7, 1, 8]$.

Given a sequence $A = [a_1, ..., a_{n-2}]$ we can reverse the procedure as follows: Let $a_{n-1} = n$. For $i = 1, 2, ..., n-1$, let $b_i$ be the vertex with the smallest label *not* in the set

$$\{a_i, a_{i+1}, ..., a_{n-1}\} \cup \{b_1, b_2, ..., b_{i-1}\}. \tag{1}$$

**Show that every tree has a leaf and that $\{(a_i, b_i) : i = 1, ..., n-1\}$ form the edges of a tree on the vertices $1, ..., n$.** *Remember to cite external resources if you use them.*

*Solution.* First, to show that the every tree has a leaf, suppose not. Since the tree is connected, the degree of every vertex is positive. Since there are no leaves, the degree of every node must be at least two. From the tutorial, we know that every graph whose nodes have degree $\geq 2$ contains a cycle. Note that trees *cannot* contain cycles so this is a contradiction.

Next to show that the sets $E = \{(a_i, b_i) : i \in [n-1]\}$ form the edges of a tree on the vertices $1, ..., n$, we show that

(a) The resulting graph is connected. We induct on the number of nodes in $G$. Let $P(n)$ be the predicate: for $G = ([n], E)$, $G$ is connected.

i. Base case. When $n = 0$, or $n = 1$, $E$ is empty so $G$ is connected. When $n = 2$, $G$ consists of the single edge $(1, 2)$ so it is also connected.

ii. Inductive step. Suppose for some constant $k \geq 2$, $P(0) \wedge \cdots \wedge P(k)$ is true. We will prove that $P(k+1)$ is true as well. Let $G = ([k+1], E)$ be a graph with edge set $E$ constructed using the above procedure (e.g. given sequence $A$, each $b_i$ is the smallest label *not* in the set described by Equation 1). Let $(a_1, b_1)$ be the first edge. Consider instead the edge set $E' = \{(a'_j, b'_j) : j \in [n-2]\}$ where $a'_j = a_j$ if $a_j < b_1$ and $a'_j = a_j - 1$ if $a_j > b_1$ and $b'_j = b_j$ if $b_j < b_1$ and $b'_j = b_j - 1$ if $b_j > b_1$ (since the labels are all distinct $j \neq i$). By the induction hypothesis, $G' = ([k], E')$ is connected. Adding back the edge $(a_1, b_1)$ and incrementing the indices greater than or equal to $b_1$ results in a connected graph: There exists a path between $b_1$ and any other node $v$ through $a_1$.

(b) The resulting graph has no cycles. Since there are $n - 1$ edges in $E$ and the graph constructed by $n$ is connected, the graph must be a tree (using the unmarked question).

The sequence $A$ is actually something called a *Prüfer sequence* which Hienz Prüfer used in 1918 to prove Cayley's formula (Claim 1).

*1 Point: explaining why every tree has a leaf. 4 Points: Showing that the graph is connected and that there are no cycles. In my solution showing the graph is connected is the difficult part so I would allot 3 points for that and 1 point for showing that the graph has no cycles, but other solutions might differ.*

b. (2 Points.) Complete the rest of the argument, why does this procedure prove Claim 1?

*Solution.* Every sequence $A$ of length $n - 2$, results in *exactly one* sequence $B$ of length $n - 2$ using the procedure described in the previous part e.g. if we started with the example sequence $A = [2, 2, 1, 1, 7, 1, 10, 10]$, *without looking at* $B$, I know its first element is 3 because *it is the vertex with smallest label not in* $A$. Further, I know that the second element of $B$ is 4 because *it is the vertex with the smallest label not in* $(A \backslash \{a_1\}) \cup \{b_1\}$ and so on.

Given a pair of sequences $A$ and $B$, there is only one way to reconstruct the tree. Thus there is "one-to-one" relationship between the sequences $A$ and labeled trees. Since there are $n^{n-2}$ possible sequences $A$, there must be the same number of labeled trees.

*1 Point: the number of lists A equals the number of labeled trees. 1 Point: explain the count of the number of lists A.*

c. (3 Points.) For tree on $n$ nodes with $n \geq 4$ how many contains exactly n-3 leaves.

*Solution.* The key observation is that every number which *does not* appear in the list $A$ (before adding $a_{n-1}$) is a leaf. To see this, note that a node $a_i$ is only added to $A$ if it is adjacent to some leaf node $b_i$ which gets added to $B$. In all iterations the neighbour of a leaf node *cannot* be a leaf node: This can only occur if there is only one edge remaining and the algorithm stops when this happens. Thus number of leaves is $n - u$ where $u$ is the number of unique indices which appear in $A$.

From there, one of two (plus the original) problem could be solved.

(a) For a tree on $n$ nodes with $n \geq 5$ how many contains exactly three leaves?

This is equivalent to computing the number of sequences of length $n - 2$ which contain *exactly* $n - 3$ *unique* elements. From the set of $n$ items, we can pick $n - 3$ to be unique and permute them. Then we pick one of the $n - 3$ to be duplicated. The duplicated item can be in one of $n-2$ positions: before the first, after the last and between any pair of adjacent elements. Since the duplicate and the original are indistinguishable, we must also divide by two. Final answer: $\binom{n}{n-3}\frac{(n-3)!(n-3)(n-2)}{2} = \frac{n!(n-3)(n-2)}{12}$.

*1 Point: observing the relationship between the elements* not *in A and the leaves. 2 Points: correct final answer.*

(b) For a tree on $n$ nodes with $n \geq 5$ how many contains exactly $n-3$ leaves?

This is equivalent to computing the number of sequences of length $n-2$ which contain *exactly three unique* elements. The number of sequences with 3 or fewer unique elements is $\binom{n}{3}3^{n-2}$. The number of sequences with 2 or fewer unique elements is $\binom{n}{2}2^{n-2}$. Each such sequence can arise from one of $\binom{n-2}{1}$ sequences counting $\leq 3$ terms. And the number of sequences with only one unique element is $n$. Each such sequence arises from one of $\binom{n-1}{2}$ sequences counting $\leq 3$ terms. Thus, the number of sequences of length $n-2$ with *exactly three unique* elements is $\binom{n}{3}3^{n-2} - \binom{n-2}{1}\binom{n}{2}2^{n-2} + \binom{n-1}{2}n$.

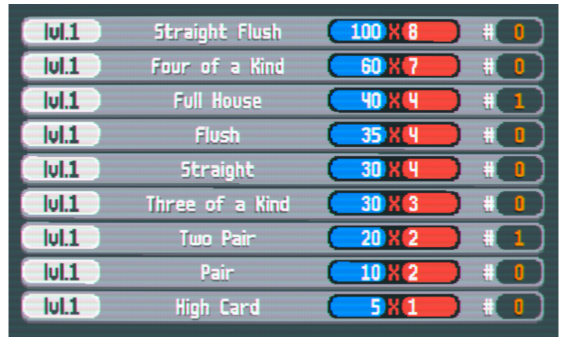
*Same as rubric as above.*

(c) Compute the expected number of leaves.

The final answer is $(n-1)^{n-2}$. I won't go into too much depth as most people probably did not solve this problem. Please read an exposition here if you are interested. Essentially it was necessary to make a further connection with the polynomial $(x_1+\cdots+x_n)^{n-2}$ and note that the monomials in the expansion correspond to labeled trees (this is not dissimilar to what we saw with the binomial coefficients which appear in $(1+x)^n$).

*The student* needs to explicitly mention that they are solving this question *if they choose to do so. There needs to be a detailed rationale for how they arrived at the solution as I am personally curious as to how someone might approach such a problem. Correct answers with a good rationale gets 2 points (1 point for observing the relationship as before). Exemplary solutions maybe shared with the class with the student's permission.*

(a) My hand (bottom): ace of spades, clubs, diamonds, 10 of spades, 6 of spades, 6 of clubs, and 3 of clubs. What's in my deck (top): 6 spades, 6 hearts, 6 clubs, and 8 hearts.




(b) Base value of the different poker hands (before adding chip value of played cards): straight flush give 100 chips and 8 multiplier, four of a kind gives 60 chips and 7 multiplier, and so on.

Figure 2: My hand (left) and base values of poker hands (right).

**Q2. [10 Points] Balatro** *(maximum 2 pages)* Balatro is a solitaire game based on poker hands. *You do not need to buy the game in order to solve the problem*[1]. The game has a bunch of jokers and card modifiers which augment the score, but *I will not be using those.*

Considering the set of cards shown in Figure 2a. I will pick five cards from the set to play and they will score according to the base values shown in Figure 2b. Scores have two components: chips (blue number shown on the left e.g. full houses are worth 40 chips) and multiplier (red number shown on the right e.g. full houses are worth 4 multiplier). Better poker hands result in more chips and (possibly) more multiplier. In addition to the base score for a hand, I also get chips for each card in the hand, added to base chip value *before* multiplying by the multiplier. Number cards are worth their face value (e.g. a nine is worth 9 chips) except aces which are worth 11 chips[2].

Let's look at an example. From the set of cards shown at the bottom of Figure 2a, I can make a full house consisting of three aces and two sixs. The base score of a full house is 40 chips with a $4\times$ multiplier. The value of the cards in the hand is 11 chips for each ace and 6 for each 6. This means that the total chip value of my hand is $40 + 3(11) + 2(6) = 85$. Together with the multiplier, the score of my hand is $85 \times 4 = 340$. Unfortunately, the *blind* (the target score) is 350.

I have one *discard* left where I can pick $n \leq 5$ cards from the set, remove them, and replace them with $n$ cards drawn uniformly from the ones remaining in my deck (shown at the top of Figure 2a). ***The goal is to determine which cards to discard in order to have the best chance of hitting the blind.*** We do so by considering the types of hands to play. Define "*successfully scoring*" as obtaining a hand worth more than 350 chips. *In the following, no proofs are required, but you need to consider all the cases. You may check your work programmatically, but make sure to state the cases and explain your calculation in your submission.*

*1 Point for each correct calculation in a group (e.g. all diamond flushes together make one point) up to a maximum of 10.*

[1] It might be helpful to watch a video of someone playing to get a sense of the game (look online for these).

[2] Face cards are worth 10 chips each, but there are no face cards so don't worry about them.

a. What is the probability that after one discard I *successfully score* with a full house?

In the following the denominator of the probabilities will depend on the number of cards discarded. If five cards are discarded then it is $\binom{26}{5}$, if four cards then it is $\binom{26}{4}$, and if three cards then it is $\binom{26}{3}$. In the following we will only consider the numerator.

*Total of 6 cases.*

(a) Three $A$, Two 10 (discard $6, 6, 3$): I want to draw one or two 10 (but not three 10) and no $A$. The number of ways this can happen is $\binom{3}{1} \cdot \binom{22}{2} + \binom{3}{2} \cdot \binom{22}{1}$.

(b) Three $A$, Two 9 (discard $10, 6, 6, 3$): I want to draw two or three 9 (but not four 9) and fewer than two 10s. Again, I can't draw any $A$. The number of ways this can happen is:

$$\binom{4}{2} \cdot \binom{3}{1} \cdot \binom{18}{1} + \binom{4}{2} \cdot \binom{18}{2} + \binom{4}{3} \cdot \binom{3}{1} + \binom{4}{3} \cdot \binom{18}{1}.$$

(c) Three $A$, Two 8 (discard $10, 6, 6, 3$): I want to draw two or three 8 (but not four 8) and fewer than two 10s and fewer than two 9s. Again, I can't draw any $A$.

$$\binom{4}{2} \cdot \binom{3}{1} \cdot \binom{4}{1} + \binom{4}{2} \cdot \binom{3}{1} \cdot \binom{14}{1} + \binom{4}{2} \cdot \binom{4}{1} \cdot \binom{14}{1} + \binom{4}{2} \cdot \binom{14}{2} \qquad \text{two 8}$$

$$+ \binom{4}{3} \cdot \binom{3}{1} + \binom{4}{3} \cdot \binom{4}{1} + \binom{4}{3} \cdot \binom{14}{1} \qquad \text{three 8}$$

(d) Three 10, Two $A$ (discard $A, 6, 6, 3$): I want to draw two (but not three) 10 and no $A$. The number of ways this can happen is $\binom{3}{2} \cdot \binom{22}{2}$.

(e) Three 9, Two $A$ (discard $A, 10, 6, 6, 3$): I want to draw three out of the four 9 and no $A$. The number of ways this can happen is $\binom{4}{3} \cdot \binom{21}{2}$

(f) Three 10, Two 9 (discard any two $A$, $6, 6, 3$): I want to draw two 10, and two or three 9 and not draw an $A$. The number of ways this can happen is $\binom{3}{2} \cdot \binom{4}{2} \cdot \binom{18}{1} + \binom{3}{2} \cdot \binom{4}{3}$.

b. What is the probability that after one discard I *successfully score* with a four of a kind?

*Total of 6 cases.*

(a) Four $A$ (discard $10, 6, 6, 3$): I want to draw the remaining $A$. There are $\binom{25}{3}$ ways this could happen.

(b) Four 10 (discard two aces, $6, 6, 3$): I want to draw all three 10. There are $\binom{23}{2}$ ways this can happen. Four 3 is calculated similarly.

(c) Four 9 (discard two aces, $6, 6, 3$): I want to draw all four 9. There are $\binom{22}{1}$ ways this can happen. Four 8 and four 7 are calculated similarly.

c. What is the probability that after one discard I *successfully score* with a straight flush?

*Total of 7 cases.*

(a) Spade straight flush $10, 9, 8, 7, 6$ (discard all $A$s, 6 of clubs, 3 of clubs): I want to draw $9, 8, 7$ of spades. There are $\binom{23}{2}$ ways this could happen.

(b) Club straight flush 10 down to 6 and club straight flush 9 down to 5 (discard 3, 6 of spades, 10 of spades, and any two aces): for the former, I want to draw $10, 9, 8, 7$ of clubs. There are $\binom{22}{1}$ ways to do this. For the latter, I want to draw $9, 8, 7, 5$ of clubs, but I do not want to draw the 10 of clubs. There are $\binom{21}{1}$ ways to do this.

(c) Diamond straight flushes 10 down to 6, 9 down to 5, 8 down to 4, and 7 down to 3 (discard 3, 6 of spades, 10 of spades, and any two aces): In all these cases, I have none of the necessary cards so I would need to draw them all. There is only one way any of these can happen.