Solution Set 0

Due Tuesday, September 4th

This homework is intended to be a self-administered placement quiz, to help you (and me) determine if you have the background for the course. You may use textbooks, your own class notes from previous courses, and material you find on the internet. Please include a list of the reference materials you used when you hand in your assignment. You may not refer to previous homework assignments and/or solution sets from the core courses from prior years or discuss this assignment with anyone other than Dannie Durand.

Please provide hand-written answers to all questions. On each question, give your reasoning and show all intermediate steps leading to your solution. Attach additional pages, if needed. Return this assignment in class on Tuesday, September 4th.

1. To solve this problem you will need a table of the genetic code.

The following DNA sequence is a fragment from the human gene encoding polypyrimidine tract-binding protein 1.

AACCA CCACC TGCGG GTCTC CTTCT CCAAG TCCAC CATCT AGGG

(a) This fragment contains the end of the coding sequence. Circle the stop codon.

AACCA CCACC TGCGG GTCTC CTTCT CCAAG TCCAC CATCT TAGGG

(b) Give the translation of the peptide encoded by this fragment.

Asn His His Leu Arg Val Ser Phe Ser Lys Ser Thr Ile

(c) For each of the following mutations, describe how the peptide would be changed by the mutation or give the modified peptide sequence.

i. The C in position 24 is mutated via a transition. TTC −> TTT. The amino acid (Phe) is unchanged.

ii. The C in position 6 is mutated via a transversion. CAC −> CAG or CAC −> CAA. In either case, the second amino acid is changed from His to Gln.
iii. The A in position 41 is mutated via a transition.

\[ TAG \rightarrow TGG. \text{ The stop codon becomes a Trp, resulting in a longer amino acid sequence.} \]

iv. A T is inserted before A in position 28.

\[ This \text{ frame shift mutation introduces a premature stop codon.} \]
2. Phenylketonuria (PKU) is a recessive disease caused by mutations in the phenylalanine hydroxylase gene on chromosome 12 in humans. In PKU, phenylalanine is not broken down and accumulates in the blood. Phenylalanine is toxic to the brain. Untreated individuals with PKU show progressive developmental delay in the first year of life, mental retardation, seizures, autistic-like behavior and a peculiar body odor. The incidence of PKU is roughly 1 in 10,000 individuals and approximately 1 individual in 50 is a carrier.

(a) A phenotypically normal woman marries a phenotypically normal man, who had a son with PKU by a previous marriage. What are the possible genotypes of the husband? The second wife? The son by the previous marriage? (Use the symbol “A” for the normal or wild type allele and “a” for the mutant.)

Husband: Aa
Second wife: Aa or AA
Son: aa

(b) Given the possible genotypes and the frequency of carriers in the population, what is the chance that the couple’s first child will have PKU?

\[
P(a \text{ from father}) = \frac{1}{2} \\
P(a \text{ from mother}) = P(\text{mother is Aa}) \times \frac{1}{2} = \frac{1}{50} \times \frac{1}{2} \\
P(\text{PKU}) = P(a \text{ from father}) \times P(a \text{ from mother}) = \frac{1}{2} \times \frac{1}{50} \times \frac{1}{2} = \frac{1}{200}
\]

(c) If the first child does have PKU, what is the probability that the second child will be normal?

\[
P(\text{PKU}) = 1 - P(a \text{ from father}) \times P(a \text{ from mother}) = 1 - \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}. \text{ Since the first child has PKU, the wife’s genotype must be Aa.}
\]

(d) If the first child has PKU, and they have 5 more children, what is the probability that only 2 of those 5 will have PKU?

The probability of a child having PKU is \( \frac{1}{4} \). Therefore, the probability of 2 children having PKU is \( \frac{1}{2} \times \frac{3}{4} \). However, there are \( \binom{5}{2} \) ways of selecting the two children. Therefore the probability that exactly two will have PKU is \( \binom{5}{2} \times \left( \frac{1}{4} \right)^2 \times \left( \frac{3}{4} \right)^3 \)

(e) If the first child has PKU, and they have 5 more children, what is the probability that at most 1 of those 5 will have PKU?

The probability of a child having PKU is \( \frac{1}{4} \). Therefore, the probability of no children having PKU is \( \frac{3}{4}^5 \), and the probability that exactly one will have PKU is \( \binom{5}{1} \times \frac{1}{4} \times \frac{3}{4}^4 \). Thus, the probability that no more than one child will have PKU is \( \frac{3}{4}^5 + 5 \times \frac{1}{4} \times \frac{3}{4}^4 \).
3. If infants with PKU are identified early via genetic screening, the impact of the disease can be lessened. Genetic tests have a false negative rate of roughly 0.02 and a false positive rate of 0.1. What is the probability that a randomly selected infant who tests positive, actually has PKU? (Hint: use Bayes’ theorem)

Let $E_D$ be the event that the patient has the disease,
$E_H$ be the event that the patient is healthy,
$E_P$ be the event that the patient’s test is positive.

\[
P(E_D|E_P) = \frac{P(E_D) \times P(E_P|E_D)}{P(E_D) \times P(E_P|E_D) + P(E_H) \times P(E_P|E_H)}
\]

\[
= \frac{\frac{1}{10000} \times 0.98}{\frac{1}{10000} \times 0.98 + \frac{9999}{10000} \times 0.1}
\]

\[
= 0.0009791384 \approx 1 \times 10^{-3}
\]
4. A clique is an undirected graph $G = (V, E)$ in which every pair of vertices is connected by an edge. If $|V| = n$, we say that $G$ is a clique of size $n$.

(a) You have a clique of size $n$. How many edges do you need to remove in order to obtain two separate cliques of size $n/2$? Give an expression in terms of $n$ and explain your reasoning.

You have $\frac{n(n-1)}{2}$ edges in the clique of size $n$ and need $2 \cdot \frac{n/2(n/2-1)}{2}$ edges in the cliques of size $n/2$.

Subtract: $\frac{n(n-1)}{2} - 2 \cdot \frac{n/2(n/2-1)}{2} = \frac{1}{4}n^2$

Alternatively, consider separating the size-$n$ clique into two parts. For each of the $\frac{n}{2}$ vertices in one part, take out the $\frac{n}{2}$ edges that connect it to the other part. By the time the parts are separated, you will have removed $(\frac{n}{2})^2 = \frac{1}{4}n^2$ edges.

(b) Suppose a connected graph, $G$, has one node of degree 4 and four nodes of degree 2. How many edges must you add to $G$ to obtain a clique of size 5? Explain your answer.

A clique of size 5 must have ten edges. The current graph has six edges. Therefore, you need to add four edges.
(c) Suppose you were writing a program that involved calculations with cliques. Which data structure would you use to represent the clique in your program, an adjacency matrix or adjacency lists? Explain the advantages and disadvantages of the data structure you chose in terms of space and running time.

Given $G = (V,E)$, where $|V| = n$ and $|E| = m$, an adjacency list uses $\Theta(n + m)$ memory. In the worst case, looking up an edge requires searching a list of size $\Theta(m)$. An adjacent matrix for this graph requires $\Theta(n^2)$ memory. Looking up an edge requires $\Theta(1)$ operations.

In a clique, each vertex is connected to $n - 1$ other vertices; in this case, $m$ is $O(n^2)$. Therefore, the space required is $O(n^2)$ in both data structures. However, each lookup is $O(1)$ in an adjacency matrix and $O(n)$ in an adjacency list. Therefore, an adjacency matrix is the appropriate data structure for representing a clique. Adjacency lists are typically used when the graph is sparse.

(d) Consider a graph with $n$ nodes and $2n$ edges. What is the minimum value of $n$ for which such a graph is possible? Explain your answer.

For a fully connected graph (i.e., a clique), the number of edges is $\frac{n(n-1)}{2}$. Therefore, $2n$ must be less than or equal to this.

\[
\frac{n(n-1)}{2} \geq 2n \\
n(n-1) \geq 4n \\
n - 1 \geq 4 \\
n \geq 5
\]
5. Let $T$ be a rooted, 3-ary tree, in which every node has either zero or three children. Let $L$ be
the number of leaves and let $N$ be the total number of nodes (leaves and internal nodes) in
$T$.

(a) Give an expression for the total number of nodes in $N$ in terms of $L$.

\[
\text{Internal nodes} : \quad \frac{(L - 1)}{2}
\]

\[
\text{Total nodes} : \quad \frac{3L - 1}{2}
\]

(b) Give an expression for $E$, the number of edges in $T$, in terms of $L$.

\[
|E| = N - 1 = \frac{3L - 3}{2}
\]

(c) What is the largest possible depth of a rooted 3-ary tree with $L$ leaves; i.e., what is
$\max_{T \in T} d(T)$?

\[
d \leq \frac{L - 1}{2}
\]

(d) Suppose we wish to represent a 3-ary tree in such a way that each leaf node contains
three null links as children. How many null links are there in a rooted, 3-ary tree with
$N$ nodes? Give your answer in terms of $N$.

\[
\text{In a full 3-ary tree, } N = \frac{3L - 1}{2}. \quad \text{Therefore the number of null links is}
\]

\[
3L = 2N + 1.
\]

\[
\text{If you calculated the number of null links under the assumption that each leaf node has}
\text{two null links, then the number of null links is}
\]

\[
2L = \frac{4N + 2}{3}.
\]