EECS 730 Background Survey

Open book, open notes, open to internet resources, but:
NO copy and paste. Develop your own solutions.
NO discussion among classmates (treat this as a take home exam)

Total point: 100
Assigned: Aug 20th, Due: Aug 27th before class meeting time
Problem 1 (25 points)
Given the following graph, show the result of different graph search:

(1) Starting at node 2, show the sequence of visited nodes using breadth-first search

(2) Starting at node 2, show a sequence of visited nodes using depth-first search

(3) What is the shortest distance between node 3 and node 5?
(4) Draw a spanning tree of the graph

(5) Outline two data structures that you may use to implement a tree
Problem 2 (25 points)

A string $T$ is a subsequence of a string $S$ if $T$ can be obtained by removing characters from $S$ without changing the order of remaining characters. For example, a string $T=“ABC”$ is a subsequence of the string $S=“ADBBCD”$ since we could obtain “ABC” by removing the second (D), the fourth (B), and the last character (D) of $S$ without changing the order of the remaining characters in $S$. As another example, $T=“ABC”$ is not a subsequence of the string $S=“ACBD”$ since there is no way that we can obtain $T$ from $S$ by removing characters from $S$ alone.

With the above definition, answer the following questions. Justify your answer.

(1) Whether the string $T=“CGAACT”$ is a subsequence of the string $S=“AACCTGC”$. If yes, specify how to obtain $T$ from $S$. If no, state your reason.

(2) Giving two string $S_1=“CGAACT”$, $S_2=“AACGACTGC”$, find a common subsequence where a string $T$ is a common subsequence to two string $S_1$ and $S_2$ if $T$ is a subsequence to both $S_1$ and $S_2$.

(3) Identify the longest common subsequence (the common subsequence with the maximal number of characters among all common subsequences) of $S_1$ and $S_2$ from above.
(4) Discuss how to develop an algorithm to solve the longest common subsequence problem in general. The input to the algorithm is a set of two sequences and the output from the algorithm is the common longest subsequence. Write down your pseudo code and explain how it works.

(5) Apply your algorithm to solve question (3).
Problem 3 (25 points)
Compute expectation and variance for random variables sampled from the following 6 distributions. The parameter(s) of each distribution is given and so you may use the same symbol computing the expectation and variance.

(I) Continuous distributions
- Normal $f(x, \mu, \sigma) \ x \in \mathbb{R}$
- Exponential $f(x, \lambda) \ x \in \mathbb{R}^+ \cup \{0\}$

(II) Discrete distributions
- Bernoulli distribution $f(k, p), k = \{0,1\}$
- Binomial distribution $f(k, p, n) k = \{0,1,2,\ldots,n\}$
- Poisson distribution $f(k, \lambda) k \in \mathbb{N} \cup \{0\}$
- Geometric distribution $f(k, p) = (1-p)^{k-1}p, k \in \mathbb{N}$
Problem 4 (25 points) Explain the following biological terms in a concise way (1~3 sentences in most cases)

(1) High throughput sequencing

(2) Chromosome

(3) Gene

(4) Promoter

(5) Protein

(6) Transcription

(7) Translation (in terms of protein synthesis)

(8) Population genetics

(9) Protein-protein interaction